The Casualty Actuarial Society *Forum*

Winter 2001 Edition

Including the Ratemaking Discussion Papers and Data Management/Data Quality/Data Technology Call Papers

To CAS Members:

This is the Winter 2001 Edition of the Casualty Actuarial Society *Forum*. It contains seven Ratemaking Discussion Papers, five Data Management/Data Quality/ Data Technology Call Papers, two committee reports, and three additional papers.

The Casualty Actuarial Society *Forum* is a non-refereed journal printed by the Casualty Actuarial Society. The viewpoints published herein do not necessarily reflect those of the Casualty Actuarial Society.

The CAS *Forum* is edited by the CAS Committee for the Casualty Actuarial Society *Forum*. Members of the committee invite all interested persons to submit papers on topics of interest to the actuarial community. Articles need not be written by a member of the CAS, but the paper's content must be relevant to the interests of the CAS membership. Members of the Committee for the Casualty Actuarial Society *Forum* request that the following procedures be followed when submitting an article for publication in the *Forum*:

1. Authors should submit a camera-ready original paper and two copies.
2. Authors should not number their pages.
3. All exhibits, tables, charts, and graphs should be in original format and camera-ready.
4. Authors should avoid using gray-shaded graphs, tables, or exhibits. Text and exhibits should be in solid black and white.
5. Authors should submit an electronic file of their paper using a popular word processing software (e.g., Microsoft Word and WordPerfect) for inclusion on the CAS Web Site.

The CAS *Forum* is printed periodically based on the number of call paper programs and articles submitted. The committee publishes two to four editions during each calendar year.

All comments or questions may be directed to the Committee for the Casualty Actuarial Society *Forum*.

Sincerely,

Dennis L. Lange, CAS *Forum* Chairperson

The 2001 CAS Ratemaking Discussion Papers and Data Management/Data Quality/Data Technology Call Papers Presented at the 2001 Ratemaking Seminar March 12-13, 2001 The Mirage Las Vegas, Nevada

The Winter 2001 Edition of the CAS *Forum* is a cooperative effort between the CAS *Forum* Committee and two CAS Research and Development Committees: the Committee on Ratemaking and the Committee on Management Data and Information.

The CAS Committee on Ratemaking presents for discussion seven papers prepared in response to its Call for 2001 Ratemaking Discussion Papers. In addition, the Committee on Management Data and Information presents three papers submitted in response to the 2001 Call for Data Management/Data Quality/Data Technology Papers.

This Forum includes papers that will be discussed by the authors at the 2001 CAS Seminar on Ratemaking, March 12-13, in Las Vegas, Nevada.

Ratemaking Discussion Papers

Data Management/Data Quality/Data Technology Call Papers

Committee Reports

Additional Papers

Ratemaking for Maximum Profitability

Lee M. Bowron, ACAS, MAAA and
Donnald E. Manis, FCAS, MAAA

RATEMAKING FOR MAXIMUM PROFITABILITY

Lee Bowron, ACAS, MAAA
Don Manis, FCAS, MAAA

Abstract

The goal of ratemaking methodologies is to estimate the future expected costs for a book of business. Using past experience, including both internal and external data, the actuary attempts to quantify the required premium level to achieve an acceptable profit.

However, if one looks at the rate activity in a market, it is apparent that company actions do not always follow the indications. Surprisingly, such decisions often lead to successful results. It seems that there must be something going on that is invisible to the naked eye? Do indications really mean so little? Or are there other factors, buried in the actuarial judgment of the experienced actuary but difficult to quantify?

It is the premise of this paper that such factors do indeed exist. One such factor is the effect of the rate change on market behavior. In this paper, we will describe one method for quantifying some of this effect. The methodology described will require much research to determine reasonable assumptions before it can used in practice. It is our hope that it will stimulate further discussion, research, and a move toward acceptance of dynamic economic principles in ratemaking.

RATEMAKING FOR MAXIMUM PROFITABILITY

Lee Bowron, ACAS, MAAA
Don Manis, FCAS, MAAA

Introduction

The goal of ratemaking methodologies is to estimate the future expected costs for a book of business. Using past experience, including both internal and external data, the actuary attempts to quantify the required premium level to achieve an acceptable profit.

However, if one looks at the rate activity in a market, it is apparent that company actions do not always follow the indications. It is not uncommon for a company to leave rates relatively unchanged even with large indicated increases. To the purist, such action seems illogical.

Even more surprising is the fact that such a decision often leads to successful results. It seems that there must be something going on that is invisible to the naked eye? Do indications really mean so little? Or are there other factors, buried in the actuarial judgment of the experienced actuary but difficult to quantify?

It is the premise of this paper that such factors do indeed exist. The great unknown in current actuarial methodologies is a function of the fact that the future book of business is not necessarily the same as the historical one. To the extent that the nature of the book changes, traditional methodologies are inadequate. They are using outdated data, based on policies that no longer figure in future costs.

For very moderate rate changes, this distortion may be minimal. However, large revisions may cause significant changes in the nature of a company's book. This is logical, because a company does not operate in a vacuum. One company's actions can have an effect on the actions of its competitors, and (perhaps more importantly) will have an effect on the behavior of its own policyholders. Will current customers renew? Will new business levels be affected?

Therefore, one major factor in future results is the effect of the rate change on market behavior. In this paper, we will describe one method for quantifying some of this effect. The methodology described will require much research to determine reasonable assumptions before it can be used in practice. It is our hope that it will stimulate further discussion, research, and a move toward acceptance of dynamic economic principles in ratemaking.

Review of Current Methodology

Every actuary knows the basic steps to developing rate indications. In a loss ratio method, premium is adjusted to current rate level, and trended if the exposure base is inflation-sensitive. Accident year losses are developed to ultimate, trended, and adjusted to reflect catastrophe risk. Credibility of the data is considered, and external data is used if necessary. Loss adjustment expenses are loaded by some method (often being treated the same as losses), and underwriting expenses are reflected. Of course, determination of the profit provision is an important, and often controversial, step. The projected loss ratio is then compared to the target loss ratio to determine the indicated rate level change.

Pure premiums methods use a similar procedure, except that projected indicated average premiums are compared to current average premiums to determine the indicated change.

The traditional methodology contains a number of implicit assumptions. Among them are:

- The future book of business will have essentially the same characteristics as the current (or historical) one.
- Rate changes will have no effect on the actions of other companies in the market.
- The indicated rate level change (if taken) will be equal to the change in premium volume.
- Rate changes will have no effect on the company's retention or ability to write new business.
- The profit provision can be determined academically, rather than being dictated by the market.

The assumption concerning profit provisions deserves further comment. The idea of setting a regulated profit provision is a function of the insurance industry's history. For much of the 20th century, competition in the industry was limited. For roughly the first ¾ of the century, bureau rates were not uncommon. In this environment, the filed rate was the same for most (if not all of the market). Therefore, it was impossible for the market to gravitate to a profit level determined by competition. This led to a "public utility" attitude toward regulation. In that environment, a regulated profit provision is not unreasonable – the consumer needs more protection when market power is absent. However, in the current, increasingly competitive market, it makes sense for insurance markets to work like other private industries.

There are many ways that one could attempt to reflect the inadequacies of traditional ratemaking assumptions. This paper will address the profit margin as a tool for reflecting the dynamic nature of the market. By using such statistics as price elasticity, we will quantify, to some extent, the way that economic forces operate on the profit margin. As you will see, such an approach would allow us to have indications that more reasonably reflect the probable future results.

Price Elasticity

Price elasticity is defined as the percentage decline in the units of a good sold for every percentage increase in the price.[1] Let X = price elasticity then,

$0 > X > -1$ (Price inelastic)

$X <= -1$ (Price elastic)

(Note that the case where $X = -1$ is defined as "unitary elastic")[2]

Price elasticity of property casualty insurance probably varies significantly by market (state) and line. For example, the sale of worker's compensation insurance with state mandated rates is much more inelastic than competitively rated reinsurance catastrophe covers.

The Rate Indication Revisited

Traditional ratemaking models make 2 assumptions:

- Rate changes will only impact average premium and not policy counts.
- Profit provisions should be determined in advance.

The common formula is $\frac{(LR + FER)}{(1 - VER - P)} - 1 = IRLC$ (1)

Where LR = Forecast Loss/LAE ratio
FER = Fixed Expense Ratio
VER = Variable Expense Ratio
IRLC = Indicated Rate Level Change.
P = Target Underwriting Profit

In a highly elastic marketplace, this approach can fail for two reasons:

1. Companies may be more concerned with maximizing profit and/or market share than profit margin. Note this is not always the case in insurance because it is so capital intensive. However, while additional volume may require additional capital, it also implies future profits (or losses) in the present value of the renewals. Therefore, an insurance company might reasonably seek to write 10 million at a 3% margin rather than 5 million at a 5% margin. Since traditional formulas determine the profit margin in advance, volume considerations are explicitly ignored.

[1] Andre Gabor, Pricing: Concepts and Methods for Effective Marketing, University Press, Cambridge, 1977, p. 16
[2] Timothy M. Devinney, Issues in Pricing, Lexington Books, 1988, p. 10

2. Most ratemaking formulas were derived at a time when agency writing was very prevalent, with a large portion of the expenses being variable. However, the increase in direct writers means that fixed expenses are a larger percentage of expenses and must be considered carefully.

   The Internet is an even more interesting example. A policy sold through a company web site would have an even greater percentage of fixed expenses, with virtually all acquisition costs as fixed.

   With a high number of fixed costs, firms could receive significant short-term benefits to the income statement by increasing volume and spreading fixed costs over additional premium.

   Of course, not all fixed expenses are truly fixed. Ultimately, all expenses are variable expenses. This paper will not examine how best to classify expenses as "fixed" or "variable," except to note that this is an increasingly sensitive assumption in the indication process.

   The formula above assumes that the fixed expense ratio will vary with the magnitude of the rate change. Suppose an insurer has a forecast 100% loss/lae ratio, a 10% fixed expense ratio, and a 15% variable expense ratio, with a 3% underwriting profit target. The traditional indicated rate level change is +34.1%.

   This assumes that fixed expenses will fall to 7.5% of premium. However, such a large increase will not likely produce nearly a 34.1% increase in premium. It is very likely that actual premium volume will decline in such an environment. Therefore, the fixed expense ratio assumed by the formula will be far too low. The opposite case can be made with declining rate levels.

   Therefore, this formula assumes that price elasticity = 0.

   The pricing actuary must consider the elasticity of her product in order to make reasonable estimates of the impact of fixed expenses.

Another common formula used by actuaries is $\frac{(LR)}{(1 - FER - VER - P)} - 1 = IRLC$ (2)

Where LR = Loss/LAE ratio
FER = Fixed Expense Ratio
VER = Variable Expense Ratio
P = Predetermined profit margin
IRLC = Indicated Rate Level Change

This formula assumes that premium remains the same as before, since fixed expenses are the same percentage as premium.

Therefore, Premium before the rate change = Premium after the rate change

Premium after the rate change = (1 + RateChange) x (Avg Premium) x (Policy count after rate change)

Premium before the rate change = (AvgPremium) x (Policy Count before Rate Change)

Since the two expressions above are equal,

Policy count after rate change = Policy Count before Rate Change/(1 + RateChange)

Therefore, percentage change in policies = 1/(1 + RateChange) - 1

Elasticity = Percentage change in policies/Percentage change in Rates

=[1/(1 + RateChange) -1]/RateChange

= -1/(1 + RateChange)

Note that for any rate increase, -1 < Elasticity < 0, so this formula implies more elastic markets than (1). In the example provided, the elasticity is approximately -.7, as a 39% increase by formula (2) would cause a 28% drop in policy counts.

This formula implies that elasticity can never be <= -1, which is not a reasonable conclusion if the insurance market is elastic.

Pricing Theory

There is considerable debate in the economics profession whether any meaningful general theories of pricing can be formulated. Few pricing managers in any business consult economic theory when setting prices.[3] However, it is well known that firms tend to pursue market share as well as profitability.[4]

Customers tend to infer the overall level of price from those items most frequently purchased. This explains why large super stores may tend to have very competitive prices ("loss leaders") for staples such as milk.[5]

---

[3] Devinney, p. 337
[4] Devinney, p. 240
[5] Gabor, p. 170

For a multi-line insurer, that may mean pricing lines that are purchased more frequently (auto) differently than those items which are purchased less often (life insurance).

For commodity products, firms tend to use market forces more than cost based pricing. Differentiated products are typically priced on a "cost plus" basis, which is analogous to traditional rate level indications.

The Model

In this section, we introduce an alternative model which can be used for pricing insurance. This type of model could also be used as a benchmark for existing pricing decisions.

The model we will develop in this section is appropriate for a highly elastic book of business. Specifically, we are going to use non-standard auto as our example because this book is not only highly elastic, but also has very different characteristics for new and renewal business.

New business tends to be highly elastic and highly unprofitable, while renewal business tends to be less elastic (but not necessarily inelastic) and highly profitable due to significant improvements in loss and expense ratios (note: this paper only considers improvements in loss ratios).

The elasticity difference is due to the typical marketing distribution of non-standard auto. Typically, new business for non-standard auto is competitively rated against a large number of companies. Renewals may be rated again, depending on agent or consumer preference, but many will be renewed without an additional comparison of rates.

Note that the elasticity of the different types of business leads to different profit assumptions in the marketplace.

A company that followed traditional actuarial pricing models for new and renewal business would be uncompetitive on new business and very competitive on renewal business. While this would maximize profit margins on any particular risk, the overall portfolio of risks would not be as profitable as a price discriminating book. And as an economist would expect, the market is much less competitive for the less elastic part of the book.

The process for building an "ad hoc" model to employ both competitive forces and the difference in new business and renewal profitability will be to calculate the profitability for each increment of proposed rate change. While the "profit maximizing" rate change could be calculated directly, we will maximize income by inspection because this will more easily allow for stochastic simulation.

Inputs

LR = Forecast Loss/LAE ratio
VER = Variable Expense Ratio
NB% = Current New Business Percentage
RB% = Current Renewal Business Percentage (1 – NB%)
FER = Fixed Expense Ratio (at current rates)
RenBet = Renewal Betterment or the expected point differential in loss/lae ratios between new and renewal business
NE = New Business Elasticity (<=0)
RE = Renewal Elasticity (<=0)
CE = Competitive Environment or the percentage change in business given no rate change
AP = Average Premium (at current rates)
CP = Premium at current rates
RC = Proposed rate change

New Business and Renewal Mix

Let's first calculate the revised New Business/Renewal Business mix.

Let AdjPol = Policies adjusted for the competitive environment (assuming no rate change)

Then AdjPol = [CP x (1+CE)]/AP

Let RCNB% = New Business (for a given rate change)

Then RCNB% = ([RC x NE] + 1) x NB%

Similarly, RCRB% = Renewal business (for a given rate change)

and RCRB% = ([RC x RE] + 1) x RB%

Since renewal and new business must combine to equal 100%, these must be adjusted for the new distribution. Therefore,

PropNB% = Proposed New Business
PropRB% = Proposed Renewal Business

PropNB% = (RCNB%)/(RCNB% + RCRB%)
PropRB% = 1 – PropNB%

New Business and Renewal Loss/LAE ratios

One of the features of non-standard auto is a major difference in the loss costs between new and renewal business. This can cause unexpected results for a large rate increase or decrease. A significant rate increase will decrease the ratio of new business to renewals and lower the loss ratio more than the rate change would traditionally suggest. Similarly, a rate decrease can cause the opposite effect. While unprofitable new business may eventually season into a profitable renewal book, the short-term pressures on results can be significant.

Let NBLR = New Business Loss/LAE ratio after the proposed rate change

Note that

LR/(1+RC) = NB% x NBLR + (RB%) x (NBLR – RenBet)

Which simplifies to

NBLR = LR/(1+RC) + (RenBet)(RB%)

And RBLR = Renewal Business Loss/LAE ratio after the proposed rate change

= NBLR – RenBet

Proposed Income Statement

In order to simplify the analysis, we will only examine underwriting income. In our example,

Income = Premiums Earned – Incurred Losses/LAE – Fixed Expenses – Variable Expenses

Let Premiums Earned = EP, then

EP = (RCNB% + RCRB%) x (AdjPol) x (AP) x (1+RC)

Incurred Losses/LAE (IL) can be defined as

IL = EP x [(NBLR)(PropNB%) +(RBLR)(PropRB%)]

Fixed Expenses and Variable Expenses are defined simply as

FE = (FER)(CP)

VE = (VER)(EP)

Maximizing Income

Developing the expressions derived above as a function of rate changes, we find that

Income = EP – IL – FE – VE

This expression could be differentiated to show that a maximum income exists as a function of the rate change, under the condition that the second derivative is less than zero. Differentiating this expression quickly becomes unwieldy, but we can demonstrate the relationship by a numerical approach. In this approach, we will simply substitute proposed rate changes at small intervals and inspecting the result for the maximum income. Here are three examples:

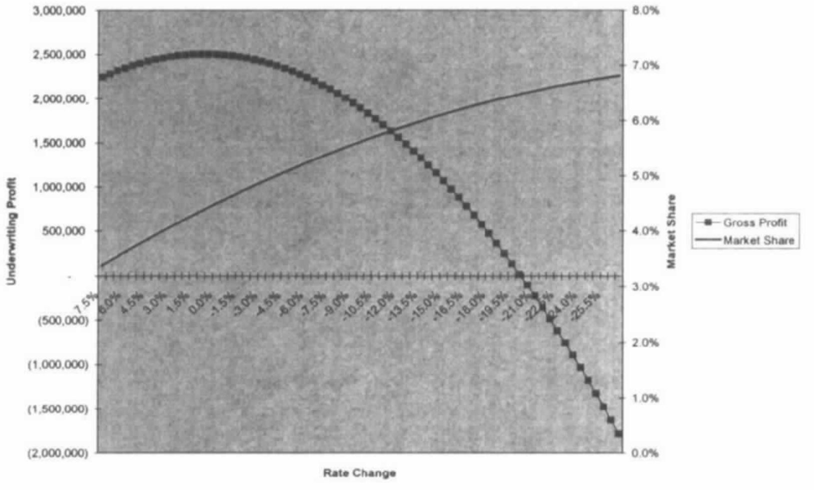
Example 1 – A low forecast loss ratio in a mature state

Assume the following inputs for the model:

LR = 60%
VER = 15%
NB% = 50%
RB% = 50%
FER = 10%
RenBet = 20%
NE = -6
RE = -2
CE = -10%
AP = $1000
CP = 20,000,000

We will also examine "proposed market share" by simply comparing the proposed premium with the market. In this example, assume the company's initial market share is 5%. The traditional indication using (1) is –12.5% with a 5% profit target. Here is a graph of profit and market share versus rate change:

Example 1: Underwriting Profit and Market Share for a Given Rate Change



The maximum profit occurs with a +1% rate change, and an underwriting profit of 2.5 million.

The earned premium, originally at 20,000,000, is projected to be at 18,000,000 for the revision with no rate change due to competitor actions.

However, a 1% increase causes a 6% decrease in New Business counts and a 2% decrease in renewal counts. Since New Business and Renewals are each half of the book, the overall counts decrease by 3.5%. Average premium increases by 1%. Therefore, the proposed earned premium (000's) is:

Current Premium	20,000
Competitive Env	-10%
Premium if No Rate Change	18,000
Policy Counts Change	-4.00%
Average Premium Change	1.00%
Premium if +1% Change	17,453

The forecast loss/alae ratio is 60%. Renewal business is projected to be 20 points better than new business. Using the formulas derived above,
NBLR = .60/1.01 + (.20)(.50) = .6941

And RBLR = .6941 - .20 = .4941

The slight increase will readjust the renewal and new business percentages. Using the formulas above, we get:

PropNB% = .4896
PropRB% = .5104

This implies that IL = 10,331,640

Since FE = (.10)(20,000,000) = 2,000,000 and VE = (.15)(17,452,800) = 2,617,920

So Income = 17,452,800 – 10,331,640 – 2,000,000 – 2,617,920 = 2,503,240

A firm may decide to trade short-term profits for additional market share and lower rates further. "Market share" can be considered a measure of long term profitability. In the example above, initial market share was 5%. Assuming that the overall market is neither growing nor decreasing, the new market share will simply be the change in premium times the initial market share. In this example, premium changes by –12.7% which decreases market share to 4.4%.

Example 2 – A high forecast loss ratio in a new state with high fixed expenses

Assume the following inputs for the model:

LR = 80%
VER = 15%
NB% = 80%
RB% = 20%
FER = 35%
RenBet = 20%
NE = -6
RE = -2
CE = -10%
AP = $1000
CP = 3,000,000

Initial Market Share = 1.0%

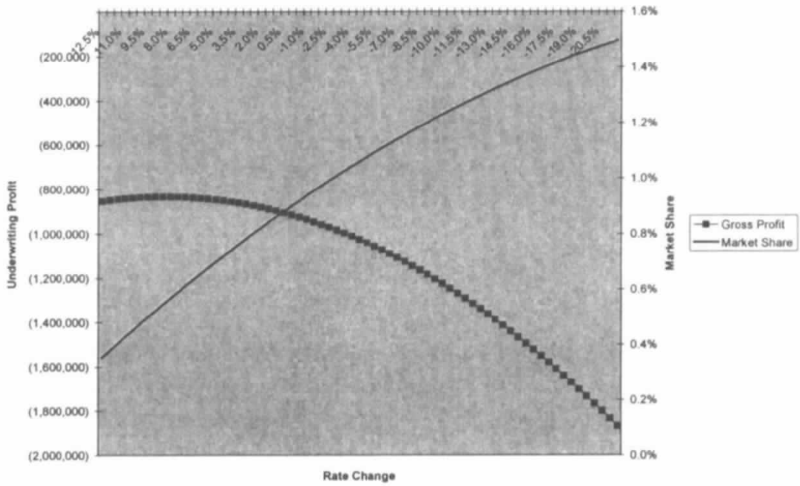
Assume this is a new state which is performing poorly. This has happened for two reasons:

- The loss ratio is above expectations
- The premium volume is well below expectations, which has caused the fixed expense ratio to be extremely high

In this case, "fixed expenses" are projected to remain high on an absolute basis over the next rate revision. This may be due to advertising contracts, leases, or other long-term commitments, but we can assume that they will not be eliminated.

The traditional formula would indicate an increase of +44%. However, the profit maximizing increase is just 8.5%. Note that in this case the maximum profit is actually a loss of 834K. However, given the fixed expenses, this is the best that can be projected.

Example 2: Underwriting Profit and Market Share for a Given Rate Change



Example 3 – A moderate loss ratio in a mature state with high fixed expenses and low variable expenses

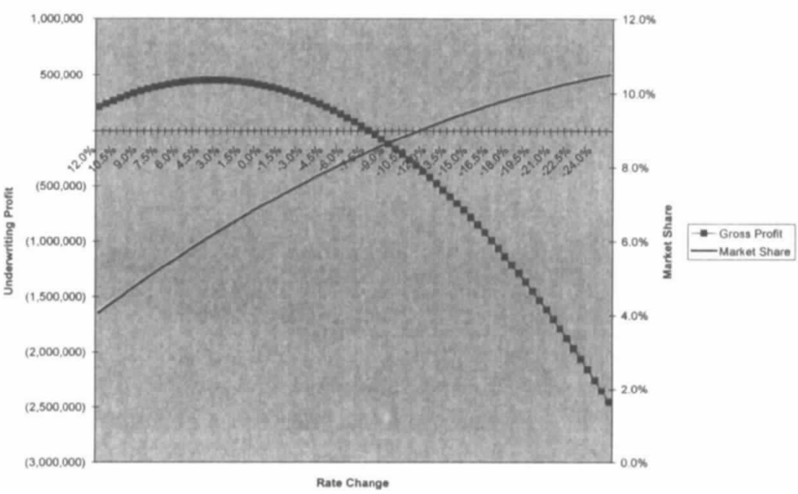
Assume the following inputs for the model:

LR = 71%
VER = 7.5%
NB% = 50%
RB% = 50%
FER = 17.5%
RenBet = 20%
NE = -6
RE = -2
CE = 0%
AP = $1000
CP = 10,000,000

Initial Market Share = 7.0%

In this example, the traditional model indicates an increase of +1.1%. This is not a dramatically different result than the +4.0% increase that the "profit maximizing" model produces.

Example 3: Underwriting Profit and Market Share for a Given Rate Change



Using the traditional formula in this case would cause a decline in underwriting income of only (25K) and gain in premium of 1484K. Therefore, one should probably look closely at market share in such a situation.

Introducing Simulation

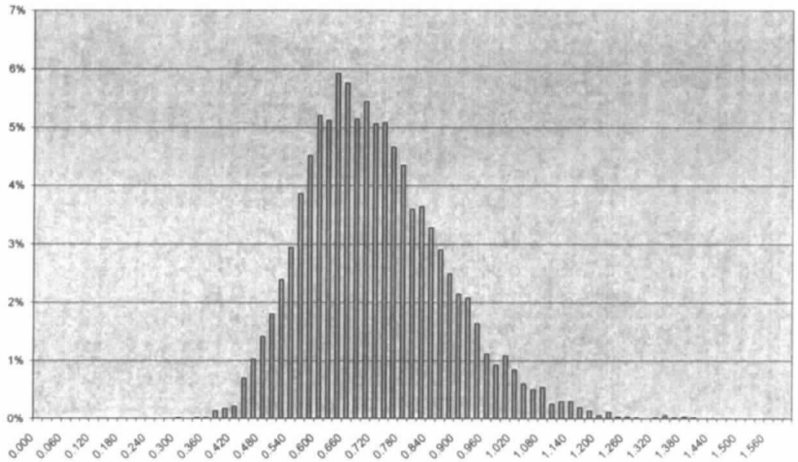
How does the "profit maximizing" model respond to differences in expectations from the traditional model?

5000 simulations of results were run of Example 3, with distributions substituted for the loss ratio, new business elasticity, and renewal elasticity.

The most sensitive variable in a rate indication is the forecast loss ratio.
We replaced the loss ratio pick in the example above of 71% with a lognormal distribution with a mean of .71 and a standard deviation of .15. We also truncated the result to a minimum of 0.

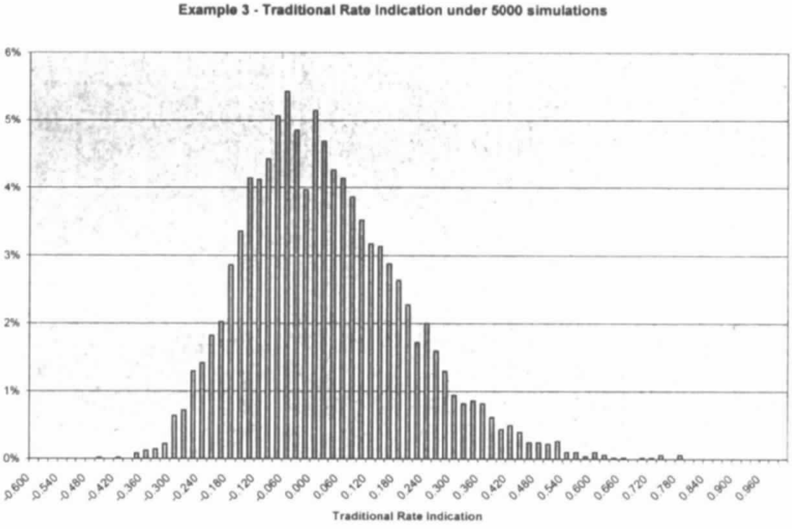
The variance was selected judgmentally, but a review of the histogram of the results shows a reasonable approximation of results for a line not subject to significant catastrophe losses:

Example 3 - Loss ratio of 5000 simulations (Mean = .71, Std. Dev. = .15)

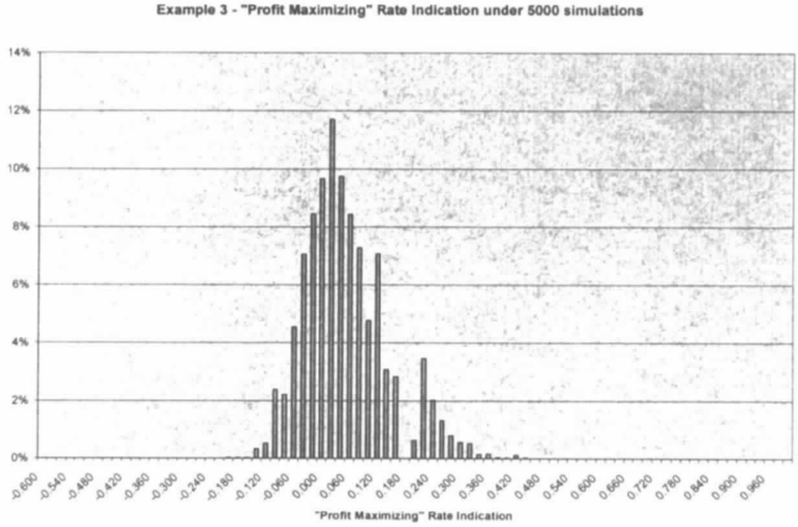


Ideally, empirical studies would determine the elasticity of the product. Also, the elasticity would be expected to change somewhat over the range of rate increases. For example, the elasticity of a 10% rate increase would probably be different than a 1% increase. However, we have not changed our elasticities in this example.

Not surprisingly, "traditional indications" correlate closely with the loss ratio, since elasticities do not impact the result of this formula:



The "profit maximizing" indications, shown below, provide an interesting contrast to the traditional indications:



Note that the shape of the distribution is remarkably different from the traditional indication. The standard deviation of the results from the traditional indication is .171 versus .092 from the "profit maximizing" model.

Conclusion

There are many assumptions in our examples that can be refined and improved. For example, price elasticity is not constant for all indications – as a result our results would be less accurate for extreme rate indications. The emphasis of the paper is on a method of thinking about rate indications in a dynamic market. As we have shown, traditional methodologies do not adequately account for the effect rate changes have on retention and other economic factors.

We hope that our paper will lead to further research. For the model to be usable in the real world, empirical studies will need to develop reasonable assumptions for price elasticity functions, distributions of new versus renewal business, and other model inputs. There are also several simplifying assumptions which would need to be refined.

This proposed approach is only a first step, but we are convinced that it is a step in the right direction. Companies that are able to reflect market forces in their rate analysis can gain a competitive advantage. A ratemaking approach that considers price elasticity to maximize profit would be a useful tool by itself, but could be even more valuable as a component of dynamic financial analysis.

As stated in our introduction, these types of concerns are reflected indirectly every time an actuary chooses not to propose the indicated rate level. With this approach, an insurance company has a better chance of measuring the effect of such decisions, creating a rate structure that balances profit and market concerns.

Considerations in Estimating Loss Cost Trends

Kurt S. Dickmann, FCAS, MAAA and
James R. Merz, FCAS, MAAA

Abstract

The application of loss trends has long been a fundamental part of the ratemaking process. Despite this, the actuarial literature is somewhat lacking in the description of methods by which one can estimate the proper loss trend from empirical data. Linear or exponential least squares regression is widely used in this regard. However, there are problems with the use of least squares regression when applied to insurance loss data.

In this paper, some common pitfalls of least squares regression, as it is commonly applied to insured loss data, and two alternative methods of evaluating loss trends will be illustrated. Both methods are based on simple least squares regression, but include modifications designed to account for the characteristics of insurance loss data.

The results of various methods are compared using industry loss data. Stochastic simulation is also used as a means of evaluating various trend estimation methods.

The concepts presented are not new. They are presented here in the context of analyzing insured loss data to provide actuaries with additional tools for estimating loss trends.

Introduction

This paper is organized into eight sections. The first section will describe the importance of estimating loss cost trends in Property/Casualty ratemaking. In addition, it will introduce the common industry practices used to estimate the underlying loss cost inflation rate.

The second section will provide a review of basic regression analysis since regression is commonly utilized for estimating loss trends. It will also describe other relevant statistical formulae.

The third section will describe some characteristics of insured loss data. This section will describe how insured losses violate some of the basic assumptions of the ordinary least squares model. It will also describe the complications that result because of these violations.

The fourth section will describe several methods that can be utilized along with informed judgement to identify outliers.

The fifth and sixth sections will describe two alternative methods that address the shortcomings of ordinary least squares regression on insured loss data.

The seventh section applies the common method of exponential least squares regression and the two alternative methods to industry loss data and compares the results.

In the last section, the performance of exponential least squares regression and the alternative methods will be evaluated using stochastic simulation of loss data with a known underlying trend.

While the determination and use of credibility is an essential component of loss trend determination, it is beyond the scope of this paper. However, the concepts and methods presented here apply equally to the determination of the trend assigned to the complement of credibility. The methods presented here are designed to extract as much information about the underlying trend from the available data. They are not intended to minimize the importance or use of credibility.

In addition to credibility, there are many other considerations that must be taken into account when applying loss trends, such as the effect of limits and deductibles. These issues are beyond the scope of this paper.

Section 1: Actuarial Literature and Industry Practice

In the ratemaking process, it is widely agreed that trend selection is the component that requires the most judgement.[1] According to the *Actuarial Standards of Practice,* the application of the appropriate trending procedures is essential to estimating future costs in the determination of rates.[2]

Despite the importance of trending in ratemaking and the degree of judgment required, there is little written specifically regarding the determination of loss trends. Most ratemaking papers cite trending as an integral part of the process and describe the author's selected approach. This is entirely appropriate as the subject of these papers is ratemaking and not specifically trend estimation.

The actuarial literature is sparse on the process of selecting the type of data to evaluate, preparing trend data, choosing the most appropriate model and assessing the appropriateness of the selected trends.

There are papers addressing several of the important basic issues of trending. These include the appropriate trending period and the overlap fallacy.[3] In addition, the CAS examination syllabus addresses the permissibility of using calendar year data to determine trends applied to accident year data.[4] These authors have well and fully addressed these topics and they need not be revisited.

---

[1] David R. Chernick, "Private Passenger Auto – Physical Damage Ratemaking", p. 6.

[2] ASP #13...

[3] Chernick, ibid., Charles F. Cook, "Trend and Loss Development Factors", CAS Proceedings, Vol. LVII, p. 1 and McClenahan, Foundations of Casualty Actuarial Science, 2d. Ed., Casualty Actuarial Society, Arlington, VA, 1990, Chapter 2.

[4] Cook, ibid.

In much of the syllabus material, both past and present, there are considerable differences between the types of data used for trending and the amount of discussion dedicated to the selection of the trend. Generally, each paper selects either calendar or accident year data and utilizes either the simple linear or exponential regression model with little guidance regarding which is more appropriate or discussion of the data to which the model is applied. These omissions are understandable since the subject of the articles is ratemaking, of which trend selection is only one component. There are acknowledgements of a need for better loss trending procedures contained in several papers.

A survey of rate filings was conducted to assess common industry practice. From this review, it is difficult to know definitively the amount of analysis that underlies the selection of trends. However, each company and the one rating agency examined display four-quarter-ending calendar year data with either simple linear or exponential regression results to support loss trend selections.[5]

As illustrated in both literature and practice, it is common in the Property & Casualty industry to estimate loss cost trends using either linear or exponential least squares regression. This is understandable since least squares regression is familiar to both regulators and company management. Further, least squares regression has been integrated into all commonly used electronic spreadsheet packages.

The validity of using linear or exponential least squares regression, the basic assumptions of regression analysis and the characteristics of loss data, in evaluating ratemaking trends has not been widely addressed. When selecting a model to estimate future trends, it is important to consider whether the data used violates assumptions of the model.

Loss Data

An essential consideration in evaluating loss trend involves the selection of the type of loss statistics to analyze. It is often useful to analyze both paid and incurred loss frequency and severity if available.

---

[5] Allstate, Nationwide, Progressive, State Farm and ISO

For example, paid claim counts may include claims closed without payment. Therefore, changes in claim handling procedures during the period under review may affect the trend estimate. Likewise, changes in case reserving practices and adjuster caseloads may affect incurred and/or paid severity amounts.

Analysis of both paid and incurred amounts, or amounts net versus gross of salvage and subrogation, can assist in identifying changes in claims handling. In any event, the loss statistics used should be defined consistently throughout the experience period. For example, if the paid loss amounts are recorded gross of salvage and subrogation for a portion of the time period, and net for the remaining, the amounts should be restated to a consistent basis prior to analysis.

Section 2: Least Squares Regression Basics

Least squares regression is a general term that refers to an extensive family of analytical methods. All of these methods share a common basic form.

$$Y_i = f(\vec{X}_i, \vec{\beta}_i) + \varepsilon_i$$

where,

Y_i is the i^{th} observation of the response variable.

$\vec{\beta}_i$ is a vector of model parameters to be estimated.

$\vec{X}_i$ is a vector of the the independent variables.

ε_i is the random error term.

Regression models are designed to use empirical data to measure the relationship between one or more independent variables and a dependent variable assuming some functional relationship between the variables. The functional relationship can be linear, quadratic, logarithmic, exponential or any other form.

The important point is that the functional relationship, the model, is assumed prior to calculation of the model parameters. Incorrect selection of the model is an element of parameter risk.

In addition to selection of the model, regression analysis also involves assumptions about the probability distributions of the observed data. This is essential in the development of statistical tests regarding the parameter estimates and the performance of the selected model.

Simple Linear Regression

The most common form of regression analysis is simple linear regression. The simple linear regression model has the following form.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where,

Y_i is the i^{th} observation of the response variable.

β_0 and β_1 are the model parameters to be estimated.

X_i is the i^{th} value of the independent variable.

ε_i is the random error term.

The parameters of the regression model are estimated from observed data using the method of least squares. This method will not be described in detail here. It is sufficient for our purpose to note that the least squares estimators, b_i, have the following characteristics:

1. They are unbiased. That is, $E[b_i] = \beta_i$.
2. They are efficient. The least squares estimators have the minimum variance among all unbiased linear estimators.
3. The least squares estimators are the same as the maximum likelihood estimators when the distributions of the error terms are assumed to be independent and normally distributed with a mean of zero and a variance of σ^2.

Because the normal distribution of the error terms is assumed, various statistical inferences can be made. Hypothesis testing can be performed. For example, the hypothesis that the trend is zero can be tested. Confidence intervals for the regression parameters can be calculated. Also, confidence intervals for $\hat{Y}$ and a confidence band for the regression line can be calculated. These very useful results make simple linear regression appealing.

Exponential Regression

While linear regression models are often satisfactory in many circumstances, there are situations where non-linear models seem more appropriate. Loss cost inflation is often assumed to be exponential. The exponential model assumes a constant percentage increase over time rather than a constant dollar increase for each time period.

The general form of the exponential regression model is given by

$$Y_t = \gamma_0 + \gamma_1 e^{\gamma_2 X_t} + \varepsilon_t$$

The parameter estimates of a non-linear regression model usually cannot be described in closed form. Therefore, numerical methods are used to determine parameter estimates using either the least squares or maximum likelihood method. Often electronic spreadsheet software will include tools to estimate the parameters for several non-linear regression models.

As with linear regression, statistical inferences such as confidence intervals for the parameter estimates, hypothesis testing and a confidence band for the fitted curve can be made.

The Exponential to Linear Transformation

In practice, the linear regression algorithm is often applied to the natural logarithm of the observed data. This transformation of the observed data simplifies the calculation of the regression parameters. However, in using this approach the analyst has, perhaps unknowingly, assumed the error terms are lognormally distributed rather than normally distributed.

The observed data is modeled using the equation,

$Ln(Y_t) = \beta_0 + \beta_1 X + \varepsilon_t$

This transformation is equivalent to the model,

$Y_t = Ke^{\beta_1 X_t} \cdot \left[e^{\varepsilon_t}\right]$, where $K = e^{\beta_0}$ and e^{ε_t} is the error term.

and the trend is obtained from the linear least squared regression estimate of β_1.

If the error term of the linear regression model, ε_i, is assumed to have a N(0,σ) distribution, it can be shown that the error term in the transformed model is lognormal with expected value $e^{\sigma^2/2}$. The error terms are positively skewed. This distribution of the error terms in the linearized model may be preferable to the normal distribution if the analyst believes it is more likely that observed values are above the mean than below the mean. This certainly may be the case with insured loss data.

Note that the lognormal distribution of the error term in the linearized model affects the calculation of confidence intervals and test statistics for the model. The familiar forms of the test statistics based on the normal distribution do not apply.

The Coefficient of Determination, R^2

Perhaps the most cited statistic derived from regression analysis is the coefficient of determination, R^2. R^2 can be interpreted as the reduction of total variation about the mean that is explained by the selected model. When R^2 is closer to one, the greater is the modeled relationship between X and Y, whether the model is linear, exponential or some other form.

The Durbin-Watson Statistic

The Durbin-Watson statistic, D, is used to test for serial correlation of the residual errors, e_t. The value of D is calculated from the observed and fitted values of Y, where $e_t = (Y_t - \hat{Y}_t)$.

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

This value is compared to critical values, d_L and d_U, calculated by Durbin and Watson. The critical values define the lower and upper bounds of a range for which the test is inconclusive. When $D > d_U$, there is no serial correlation present. When $D < d_L$, there is some degree of serial correlation present.[6]

Section 3: Insured Loss Data

There are several distinct characteristics of insured loss data that should be recognized when selecting a regression model. In broad terms, one expects data to be comprised of an underlying trend, a seasonality component, a possible cyclical nature and a random portion.[7] These traits make the estimation of the underlying trend more difficult and the rigid use of simple linear or exponential regression imprudent.

Unusual Loss Occurrences

The nature of insured losses may violate the common assumptions of simple linear or exponential least squares regression. For example, loss events that cause widespread damage can generate extraordinarily high claim frequencies in a given time period. The reverse, a time period with an extraordinarily low claim frequency, is unlikely. A similar skewness can occur in severity data for small portfolios or, almost certainly, in medium to large portfolios of liability risks due to shock losses. Examples of these characteristics are evident in trend data provided by the Insurance Services Office.

Widespread Loss Events

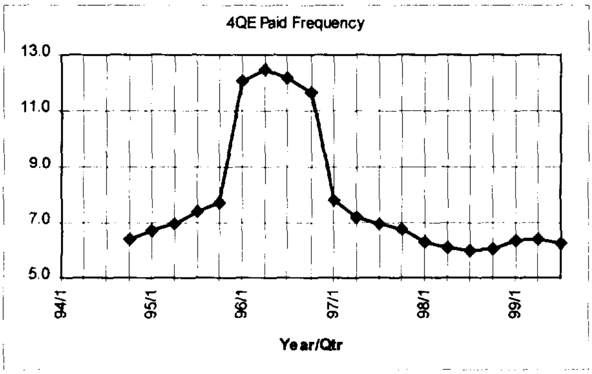
In the chart below of Homeowner claim frequencies as reported by the Insurance Services Office for the state of Oregon, there is an obviously unusual occurrence in the first quarter of 1996. The increase in claim frequency over the prior annual period is over 50%.

---

[6] Neter, et. al., Applied Linear Statistical Models, 4th ed., McGraw-Hill, Boston, 1996 p. 504.

[7] Spyros Makridakis and Steven C. Wheelwright, Forecasting Methods for Management, 5th Ed., John Wiley & Sons, New York, 1989, p. 96.

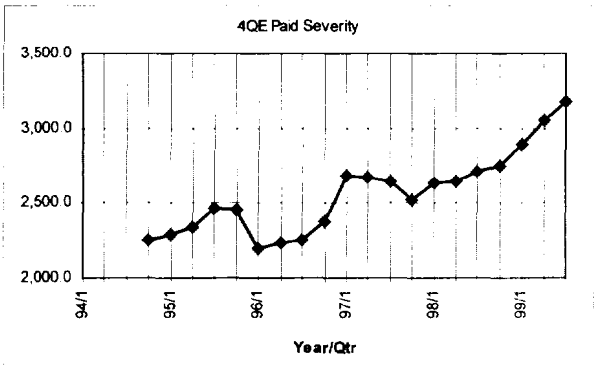
Oregon Homeowners



Because the data is twelve-month-moving, the dramatic rise in frequency that occurred in the first quarter of 1996 is transferred to the subsequent three observations. Therefore, the error terms are not independently distributed, as commonly assumed, due to the construction of the data.

A review of the severity data for the same time period shows a corresponding, though less dramatic, drop in claim severity. This is typical of a high frequency, low severity weather loss event. This drop in claim severity may go unnoticed if it were not for the associated increase in frequency. Again, due to the twelve-month moving organization of the data, the error terms are not independently distributed.

Oregon Homeowners

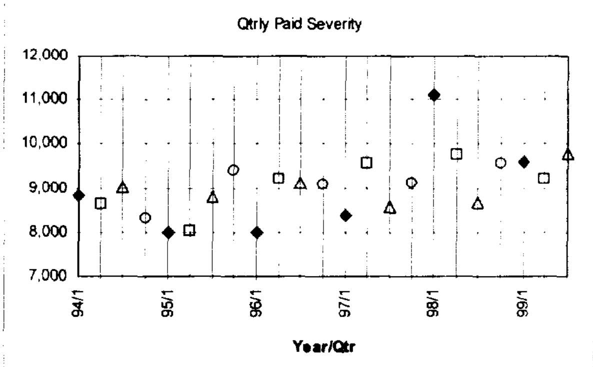


Shock Losses

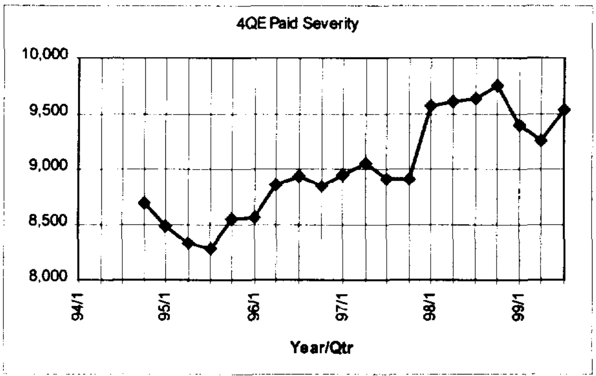
A high severity claim in a small portfolio may cause a distortion in the data and affect the trend calculated by ordinary least squares methods if no adjustments are made. A visual inspection of Nevada Private Passenger Auto Bodily Injury severity data provided by the Insurance Services Office shows an unusual occurrence in the first quarter of 1998.

The quarterly data shows the elevated severity in the first quarter of 1998 neatly as one high point while the four quarter ending data exhibits this phenomena as a four point plateau. This phenomenon occurs more often in smaller portfolios, even when utilizing basic limit data.

Nevada PPA – Bodily Injury Liability



Nevada PPA - Bodily Injury Liability



Effects of Unusual Loss Occurrences

While the cause of these events is dissimilar, the result on the data is the same. One may expect the distribution of the error term for claim frequency and severity to be positively skewed, rather than normally distributed as commonly assumed. The lognormally distributed error terms of the transformed exponential regression model may be more appropriate than the exponential model with normally distributed errors.

As demonstrated above, insured loss frequency and severity data may exhibit abnormally high random error. If these errors occur early in the time series, the resulting trend estimates from least squares regression will be understated. Conversely, if the shock value occurs late in the time series, the trend estimate will be overstated. The use of twelve-month-moving data compounds this effect since the shock is propagated to three additional data points.

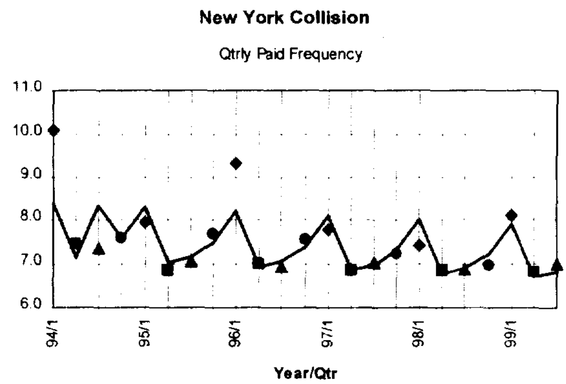
There are several methods available to identify outliers and measure their influence on the regression results. These include Studentized Deleted

Residuals, DFFITS, Cook's Distance and DFBETAS.[8] The identification of such occurrences is addressed in section four below.

Seasonality of Data

The nature of insurance coverage creates seasonal variation in claim frequency and severity. For example, winter driving conditions may cause higher Collision and Property Damage Liability claims in the first quarter. Similarly, lightning claims may be more prevalent during the summer months in certain states. The probability of severe house fires may be higher during the winter months. Auto thefts may be more frequent in summer months causing elevated severity for Comprehensive coverage.

When reviewing New York Private Passenger Auto data for Collision coverage on a quarterly basis, one can see the seasonal nature of claim frequencies. This seasonality can be illustrated by grouping like quarters together.



Generally, the use of twelve-month-moving data is a convenient method for adjusting the seasonal nature of insured losses. However, four-quarter-ending

[8] Neter, et. al., ibid., and Edmund S. Scanlon, "Residuals and Influence in Regression", CAS Proceedings, Vol. LXXXI, p. 123.

data creates serially correlated errors when used in ordinary least squares regression.

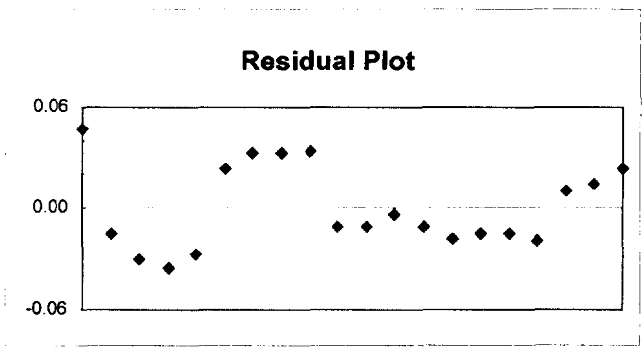
Serially Correlated Error

Actuarial literature shows trend data organized in a variety of ways. Some authors use twelve-month-moving calendar year data observed quarterly, others use accident year data observed annually, still others use calendar quarter data observed quarterly. Each format has advantages and disadvantages. It is important to recognize the implications of the data organization on the regression results.

Any organization of data that has overlapping time periods from one point to the next, by its construction, results in serially correlated error terms. Serial correlation of error terms occurs when the residual errors are not independent. This result is shown for twelve-month-moving calendar year data in Exhibit 2 using the Durbin-Watson statistic.

Additionally, one can plot residuals to detect serial correlation. Below the residual plot is displayed for twelve-month-moving New York Collision frequency. As one can see, the errors for adjacent points are related. As noted above, the independence of the error terms in ordinary least squares regression is generally assumed and certain conclusions about the regression statistics are based on this assumption.

New York Collision Frequency



According to Neter, et. al., when this assumption is not met the following consequences result.

1. The estimated regression coefficients are still unbiased, but they no longer have the minimum variance property and may be quite inefficient.
2. Minimum Squared Error (MSE) may seriously underestimate the variance of the error terms.
3. The standard deviation of the coefficients calculated according to ordinary least squares procedures may seriously underestimate the true standard deviation of the estimated regression coefficient.
4. Confidence intervals and tests using the t and F distributions are not strictly applicable.

Remedial Measures

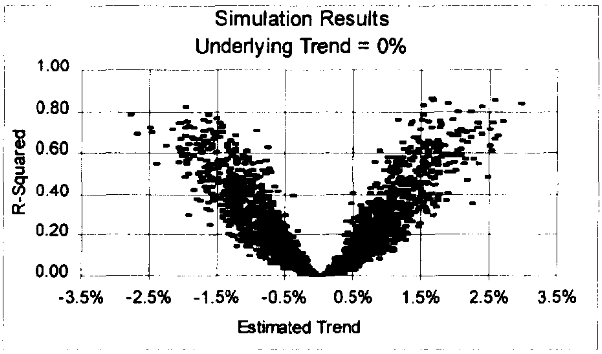
Each of the first two issues with the insured loss data, widespread loss events and extraordinary claim payments, can be resolved by removing outlying points before calculating the exponential or linear regression. The removal technique must rely on statistical tests and actuarial judgment. This will be discussed in the following section. Seasonality and serial correlation can be addressed using regression with indicator variables on quarterly data. Regression with indicator variables explicitly incorporates seasonality as a component of the model. The use of quarterly data eliminates the serial correlation resulting from the use of overlapping time periods.

Comments on Goodness-of-Fit

Estimating the underlying trend in a given dataset entails more than simply fitting a line to a set of data. During the estimation process, it is important to determine whether the underlying assumptions are met and whether the equation accurately models the observed data.[9]

Many consider R^2, the coefficient of determination, the most important statistic for evaluating the goodness-of-fit. The coefficient of determination is the proportion of the data's variability over time that is explained by the fitted curve. However, it is widely agreed that this is not sufficient.[10] The coefficient of determination, by itself, is a poor measure of goodness-of-fit.[11]

To assume that a low R^2 implies a poor fit is not appropriate. It has been shown that a low or zero trend, by its nature, has a low R^2 value.[12] Also, whenever the random variation is large compared to the underlying trend the R^2 will not be sufficient to determine whether the fitted model is appropriate. One can illustrate the low R^2 values associated with data exhibiting no trend over time. The scatter plot below was generated from a simulation with an underlying trend of zero.



---

[9] Scanlon, ibid.

[10] D. Lee Barclay, "A Statistical Note on Trend Factors: The Meaning of R-Squared", CAS Forum, Fall 1991, p. 7, and Ross Fonticella, "The Usefulness of the R^2 Statistic", CAS Forum, Winter 1998, p. 55, and Scanlon, ibid. and Neter et. al., ibid.

[11] Barclay, ibid.

[12] Barclay, ibid.

The residuals between the actual and fitted points are highly useful for studying whether a given regression model is appropriate for the data being studied.[13] It is useful to graph the fitted data against the observed data to look for patterns.[14] A random scattering of residuals occurs when the fit is proper.[15] It is important that the error term not appear systematically biased when compared to neighboring points.

The use of the R^2 statistic or plots of the residuals may result in the decision that the model is an appropriate fit to the data. This conclusion applies to the historical period based on this analysis. Another consideration is the extrapolation of the trend model into the future. As McClenahan illustrates with the use of the 3rd degree polynomial, a perfect fit within the data period does not always result in the appropriate trend in the future.[16] Extrapolation beyond the data period should also be considered before the decision to proceed with the model is undertaken.

Section 4: Identification of Outliers

This section describes methods by which one can identify extraordinary values from observed loss data. These methods are designed to identify outliers from a dataset on which regression is to be performed. An excellent reference on these and other statistical methods is Applied Linear Statistical Models by Neter et. al.

Each of these methodologies cannot be applied without judgement. None of the methods is so robust as to produce reliable results in all circumstances.

---

[13] Neter, et. al., ibid, p. 25.

[14] Fonticello, ibid.

[15] Barclay, ibid.

[16] McClenahan, ibid.

Therefore, the selected points should always be compared to the original dataset.

The identification of the cause of the outlier is preferred. For example, if possible, the claims department should be consulted if a single large claim or if a widespread claims event, such as a catastrophe, appear to distort the data.

Visual Methods

When performing simple linear regression there are several visual methods which can result in easy identification of outlying points. Among these graphs are residual plots against the independent variable, box plots, stem-leaf plots and scatter plots[17]. While residual plots may lead to the proper inference regarding outliers, there are instances when this is more difficult. When the outlier imposes a great amount of leverage on the fitted regression line, the outlier may not be readily identifiable due to the resulting reduction of the residual.

Studentized Residuals

There are several standard methods that can be utilized to assist with the identification of outliers, each with advantages and disadvantages. The studentized residual detects outliers based on the proportional difference of the error term, e_i, and the variance of these errors. The studentized residual is defined:

$$r_i = \frac{e_i}{s\{e_i\}},$$

Where $s\{e_i\}$ is an estimate of the standard deviation of the residual. This estimate is easily calculated as $s\{e_i\} = \sqrt{MSE(1-h_{ii})}$, where h_{ii} is the diagonal element of the hat matrix $H=X(X'X)^{-1}X'$. Interestingly, $\hat{Y} = HY$ and $e = (1-H)Y$. The hat matrix will be used in future development of outlier identification for simplification of the formulae.

This method has the same disadvantage as identification of outliers using residual graphing. The variance of the errors includes the error of the i^{th}

[17] Neter, et. al, ibid.

observation. In addition, there is no statistical test from which one can base a decision regarding outliers.

Studentized Deleted Residuals

A significant improvement in identifying outliers uses the studentized deleted residual. For the i^{th} observation the deleted residual, d_i, is the difference between the i^{th} observation, Y_i, and the fitted point when the fitted curve includes all but the i^{th} observation, $\hat{Y}_{i(i)}$. By excluding the i^{th} observation one can determine the influence of the observation on the fitted function. Fortunately, the deleted residual can be computed relatively easily.

$$d_i = \frac{Y_i - \hat{Y}_i}{1 - h_{ii}} = Y_i - \hat{Y}_{i(i)} \text{ where } h_{ii} \text{ is the diagonal from } H.$$

The deleted residual, d_i, when studentized (divided by the estimated standard deviation of d_i), follows the t(n-p-1) distribution. Therefore, each studentized deleted residual can be tested using $t(1-\alpha/2n; n-p-1)$. Fortunately, the studentized deleted residuals, t_i, can be computed without performing n separate regressions. It can be shown that ,

$$t_i = \frac{e_i}{\sqrt{MSE_{(i)} \cdot (1 - h_{ii})}} = e_i \left[\frac{n-p-1}{SSE(1-h_{ii}) - e_i^2} \right]$$

DFFITS

One measure of influence is the DFFITS statistic. The DFFITS is the standardized difference between the fitted regression with all points included and with the i^{th} point omitted.

$$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} = t_i \cdot \left[\frac{h_{ii}}{1 - h_{ii}} \right]^{2}$$

This represents the number of standard deviations $\hat{Y}_i$ increases or decreases with inclusion of the i^{th} observation. Note that the DFFITS statistic is a function of the studentized deleted residual and can be computed without performing multiple regressions. Observations are considered outliers if the DFFITS is greater than one for medium datasets and $2\sqrt{p/n}$ for large datasets.

Cook's D

Another measure of influence is Cook's Distance measure, D_i. Scanlon utilizes Cook's D statistic to identify outliers.[18] Cook's D measures the influence of the i^{th} case on all fitted values.

$$D_i = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p \cdot MSE}$$

The denominator standardizes the squared difference measure of the numerator. Evaluation of Cook's D is accomplished by utilizing the F(p, n-p) distribution. A percentile value less than 10-20% shows little influence on the fitted values, while a percentile value of 50% or more indicates significant influence.

Fortunately, Cook's D can be calculated for each observation from a single regression using the following relationship.

$$D_i = \frac{e_i^2}{p \cdot MSE} \cdot \left[\frac{h_{ii}}{(1 - h_{ii})^2} \right]$$

As with all models good judgement is imperative and comparison to the original data is advised. In addition to the methods described above, one can calculate a confidence band around the fitted curve. Observations outside the confidence band are candidates for removal.

Each of these methods is designed to identify a single outlier from the remaining data. These techniques may not be sufficient to distinguish outliers when other outliers are adjacent or nearby. Each of these methods is extendable to identify multiple outliers from the remaining data. However, a discussion of these extensions is beyond the scope of this paper.

---

[18] Scanlon, ibid.

Section 5: Manual Intervention - Deletion/Smoothing of Outliers

Manual Intervention

The identification of extraordinary values is certainly a matter of judgement. In the analysis that follows, the determination of outliers is completed by use of visual inspection.

In many cases a visual review of the twelve-month-moving data can identify outliers. However, the occurrence of two ouliers within four quarters of each other can be difficult to detect using twelve-month-moving data. For this analysis the data is decomposed into the quarterly loss data shown below.

Table 1 – Quarterly Frequency – Oregon Homeowners

	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1994	6.167	5.778	6.194	7.319
1995	7.573	6.665	8.076	8.613
1996	24.861	8.456	7.006	6.555
1997	9.303	6.053	5.906	5.778
1998	7.300	5.301	5.592	5.986
1999	8.539	5.463	4.965	

The observed frequency in the first quarter of 1996 is identified as an outlier.

Treatment of Outliers

Once the outliers have been identified, one can proceed in several ways. First, the analyst may simply remove the outlying point from consideration and complete the analysis as if the observation did not occur. While this alternative may seem appealing, it does not allow for the reconstruction of twelve-month-moving data.

The second approach is to replace the outlier with the fitted point from the regression after removal of the outlier. This removes the outlier from the regression entirely, but allows reconstruction of the four-quarter-ending data.

The final approach is to replace the outlying point with the fitted point plus or minus the width of a confidence interval, as appropriate. This choice mitigates the extent to which the outlier affects the regression results, without removing the point entirely.

For simplicity, the authors have selected the first approach for comparison purposes but acknowledge that the other two procedures may be appropriate in other circumstances.

Parameter Estimation
Estimation of the underlying trend in the data is completed through exponential regression on the quarterly data, excluding the outliers, with indicator variables to recognize any seasonality.

Section 6: Qualitative Predictor Variables for Seasonality

This method of least squares regression recognizes the seasonal nature of insured losses through the use of qualitative predictor variables, or indicator variables. Indicator variables are often used when regression analysis is applied to time series data. Also, since the data used in this method is quarterly rather than twelve-month-moving, first-order autocorrelation of the error terms is not present. Hence, the issues that arise from such autocorrelation are eliminated.

The linearized form of the exponential regression model is given as

$$Ln(Y_t) = \beta_0 + \beta_1 X_t + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \varepsilon_t$$

Where,

Y_t is the dependent variable
X_t is the independent variable (time)
D_2 = 1, if second quarter, 0 otherwise
D_3 = 1, if third quarter, 0 otherwise
D_4 = 1, if fourth quarter, 0 otherwise
ε_t is the random error term

The model above can be viewed as four regression models, one for each set of quarterly data.

The exponential equivalents, without error terms, are

First Quarter: $Y_t = [e^{\beta_0}] \cdot e^{\beta_1 X_t}$

Second Quarter: $Y_t = e^{\beta_2} \cdot [e^{\beta_0}] \cdot e^{\beta_1 X_t}$

Third Quarter: $Y_t = e^{\beta_3} \cdot [e^{\beta_0}] \cdot e^{\beta_1 X_t}$

Fourth Quarter: $Y_t = e^{\beta_4} \cdot [e^{\beta_0}] \cdot e^{\beta_1 X_t}$

One can think of e^{β_1} as the trend component of the model and e^{β_2}, e^{β_3} and e^{β_4} as the seasonal adjustments to e^{β_0}.

Essentially, the assumption is that the rate of change in frequency or severity over time is constant for all quarters, but the level of frequency or severity differs by quarter. This differs from multiple regression models, which assume separate trends for each quarter. A single trend, rather than four different trends, is intuitively appealing for ratemaking applications.

Section 7: Comparison of Results

This section compares trend estimates derived from five estimation methods applied to industry data provided by The Insurance Services Office. The data is displayed in Exhibit 1. Exponential least squares regression on twelve-month-moving data, quarterly data and annual data are used as examples of common industry practice. The results from the exponential regressions will be compared to results derived from the alternative methods described above.

Detailed calculations using the Oregon Homeowners data are shown in the attached exhibits. The results in the tables below show the annual trend derived from each method and the associated R^2 value in parentheses.

Table 1 - Oregon Homeowners Frequency

	# Years of Observations			
Method	2 yr.	3 yr.	4 yr.	5 yr.
12 MM	-1.5% (.06)	-13.9% (.53)	-17.0% (.62)	-6.9% (.17)
Quarterly	-15.6% (.32)	-26.7% (.45)	-13.2% (.21)	-3.9% (.03)
Annual	--	-5.3% (.50)	-19.2% (.72)	-10.1% (.34)
Manual Adjustment	--	-6.8% (.79)	-8.4% (.58)	-2.6% (.20)
Indicator Variables	-9.4% (.91)	-22.2% (.75)	-10.9% (.48)	-2.6% (.27)

Table 2 – New York PPA Collision Frequency

	# Years of Observations			
Method	2 yr.	3 yr.	4 yr.	5 yr.
12 MM	0.3% (.04)	-1.7% (.43)	-2.2% (.61)	-1.9% (.58)
Quarterly	-0.6% (.00)	-1.6% (.07)	-2.8% (.17)	-1.7% (.10)
Annual	--	-0.6% (.14)	-2.3% (.66)	-1.2% (.37)
Manual Adjustment	--	--	-1.0% (.80)	-0.8% (.84)
Indicator Variables	1.7% (.83)	-0.6% (.80)	-2.2% (.76)	-1.2% (.74)

Table 3 – Nevada PPA Bodily Injury Severity

	# Years of Observations			
Method	2 yr.	3 yr.	4 yr.	5 yr.
12 MM	1.2% (.06)	3.0% (.52)	3.1% (.72)	3.1% (.78)
Quarterly	4.9% (.10)	4.3% (.20)	4.1% (.31)	2.7% (.25)
Annual	--	3.5% (.63)	2.8% (.71)	3.7% (.85)
Manual Adjustment	--	1.2% (.85)	1.9% (.65)	1.4% (.41)
Indicator Variables	9.4% (.57)	4.9% (.36)	4.0% (.37)	2.7% (.27)

The manual adjustment method and regression using indicator variables provide additional estimates of the underlying loss trend to assist the actuary in selecting appropriate adjustment for ratemaking.

Section 8: Evaluation of Methods Using Stochastic Simulation

In this section, a simulation is constructed to test the accuracy of each estimation method. Each of the five methods above is applied to the simulated data.

Personal auto severity data is simulated using a known underlying trend, a normally distributed random error term, a seasonal adjustment for each quarter and a shock variable to simulate a single large claim payment.

Simulation Parameter Estimation

Based on the Nevada PPA Bodily Injury severity analysis from the previous section the following simulation parameters were selected.

Table 5 – PPA Bodily Injury Severity Simulation Parameters

Trend	3.5%	$e^{\beta_1} - 1$
Severity Variance	5.048 10^{-2}	$MSE / \hat{Y}_0^2$
Base Severity	$8,700	$e^{\beta_0} = E[Y_0]$

Quarter	Seasonal Adjustment		Shock Probability	Shock Magnitude
First	1.000	--	1/23	20%
Second	1.013	e^{β_2}	1/23	20%
Third	0.987	e^{β_3}	1/23	20%
Fourth	1.03	e^{β_4}	1/23	20%

The shock probability and magnitude were chosen based on the observed data. Of the 23 observations, only one observation appeared to have an extraordinarily high severity. The magnitude of the shock is fixed at 20%. The simulation could be further modified to include a stochastic variable for the shock magnitude. Simulations for other states and lines of business would incorporate other parameter values based on observed data.

The simulation function is given by,

$$Ln(Y_i) = [\beta_0 + \beta_1 X_i + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4] \cdot \hat{W}_i + \hat{\varepsilon}$$

where,

$$\Pr[\hat{W}_i = (1 + \delta_i)] = 1/23,$$

$$\Pr[\hat{W}_i = 1.00] = 22/23,$$

and

$\hat{\varepsilon}_i$ is $N(0, \sigma^2)$

The shock value of the natural logarithm of the severity, $1+\delta_i$, corresponding to the shock value of the severity must be calculated. It can be shown that the value of δ_i is given by

$\frac{\ln(1+\alpha)}{\beta_0+\beta_1 X_i+\beta_2}$, where α is the shock value for Y.

Likewise, the error variance, σ^2, for $\ln(y_i)$ is derived from the estimated variance of $Y_i = MSE/\hat{Y}_0^2$ according to the following relationship.

$$\frac{MSE}{\hat{Y}_0^2} = e^{\sigma^2}\left(e^{\sigma^2}-1\right)$$

Simulation Results

Ten thousand simulated data sets were generated. The five estimation methods were applied to each data set.

It is important to note that the application of the manual intervention method assumed correct identification of the extraordinary observations in every simulation. In practice, identification of extraordinary values depends on judgement and statistical methods as described previously. Therefore, the comparison that follows may overstate the accuracy of the manual intervention method.

The table below summarizes the results of each regression method based on 10,000 simulations of twenty observations. Since the underlying trend in the simulation is known, accuracy is measured using the absolute difference between the estimated trend and the actual trend. The percentage of estimates above the actual trends is also shown in order to detect upward bias in the estimation method. Also, the percent of estimates within various neighborhoods of the actual trend are calculated.

The simulation was constructed with a seasonal component and outliers. Therefore, it is not surprising that the manual intervention method that excludes the outliers and includes quarterly indicator variables produces good results.

Table 5 – Comparison of Methods (based on 10,000 simulations)							
			Percentage of Estimates				
Method	Average Trend Estimate	Average Absolute Difference	Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	3.52%	0.82%	50.7%	37.7%	54.1%	66.9%	.74
Quarterly	3.33%	0.91%	44.1%	34.5%	49.0%	62.5%	.34
Annual	3.51%	0.93%	50.2%	33.6%	48.6%	61.3%	.75
Indicator Variables	3.51%	0.92%	50.4%	34.4%	49.3%	62.0%	.48
Manual Adjustment	3.50%	0.81%	49.4%	37.7%	53.9%	67.6%	.54

A similar process can be used to simulate frequency data which include the probability of loss events that produce large numbers of claims.

Other Simulation Results

Four other simulations were performed. The first compares results when no shocks are present. The second simulation included only data when shock values were present. The third simulation included shocks early in the time series only. The final simulation included shocks only late in the time series.

NO SHOCKS

Table 6 – Comparison of Methods (based on 10,000 simulations)							
			Percentage of Estimates				
Method	Average Trend Estimate	Average Absolute Difference	Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	3.50%	0.69%	50.0%	43.2%	60.7%	75.0%	.80
Quarterly	3.33%	0.78%	43.2%	39.2%	55.6%	68.9%	.40
Annual	3.51%	0.78%	50.7%	39.0%	55.4%	68.8%	.81
Indicator Variables	3.51%	0.78%	50.8%	39.2%	55.4%	68.9%	.53
Manual Adjustment	3.51%	0.78%	50.8%	39.2%	55.4%	68.9%	.53

The results of this simulation show that there is little difference between traditional regression techniques and regression using qualitative predictor variables for seasonality.

ALL SHOCKED

Table 7 – Comparison of Methods (based on 10,000 simulations)

Method	Average Trend Estimate	Average Absolute Difference	Percentage of Estimates				
			Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	3.52%	0.89%	49.9%	35.2%	50.0%	63.0%	.70
Quarterly	3.35%	0.97%	44.7%	31.9%	46.3%	58.9%	.31
Annual	3.53%	1.01%	50.5%	30.5%	44.6%	57.3%	.72
Indicator Variables	3.53%	0.98%	50.6%	31.4%	45.5%	58.6%	.45
Manual Adjustment	3.51%	0.81%	49.7%	37.9%	54.0%	67.2%	.54

The results of the simulation using only data with shocks illustrate the increased accuracy of the manual adjustment method described previously under these circumstances.

SHOCKED EARLY

Table 8 – Comparison of Methods (based on 10,000 simulations)

Method	Average Trend Estimate	Average Absolute Difference	Percentage of Estimates				
			Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	1.68%	1.88%	6.7%	12.6%	19.3%	26.4%	.35
Quarterly	1.87%	1.78%	11.9%	15.4%	23.2%	30.5%	.16
Annual	1.93%	1.77%	14.1%	15.8%	23.5%	31.6%	.46
Indicator Variables	2.05%	1.66%	15.1%	17.4%	25.5%	34.9%	.33
Manual Adjustment	3.50%	0.84%	49.9%	36.6%	52.1%	65.6%	.53

This simulation illustrates the understatement of trend estimates by traditional methods when shock values occur early in the time series. While proper elimination of the shocks may be difficult, this simulation shows the value of the proper identification.

SHOCKED LATE

Table 9 – Comparison of Methods (based on 10,000 simulations)

Method	Average Trend Estimate	Average Absolute Difference	Percentage of Estimates Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	5.37%	1.93%	93.3%	12.4%	19.1%	26.6%	.79
Quarterly	5.23%	1.85%	89.5%	15.5%	23.1%	31.3%	.39
Annual	5.66%	2.22%	93.5%	10.6%	16.8%	22.7%	.80
Indicator Variables	5.50%	2.07%	92.5%	12.1%	18.7%	25.5%	.52
Manual Adjustment	3.52%	0.85%	50.3%	37.1%	52.3%	65.1%	.54

This simulation illustrates the overstatement of trend estimates by traditional regression techniques when shocks occur late in the time series.

Conclusion

The regression concepts discussed here are not new to actuaries. Nor are the characteristics of insured loss data. Actuaries are familiar with the stochastic nature of claim frequency and severity. Actuaries are also keenly aware of the potential for loss events, be they weather events that generate an extraordinary number of "normal" sized claims, or single claims with extraordinary severity, that do not fit the assumptions of basic regression analysis.

While outlier identification techniques are described in section four, they have not been applied to the industry data. The evaluation of these techniques is a subject worthy of further research. In addition, the authors would welcome development of techniques to discriminate between random noise and seasonality, to identify turning points in the trend and to distinguish between outliers and discrete but "jumps" in the level of frequency and severity.

Hopefully, the authors have presented some additional tools for ratemaking and stimulated interest in developing trend estimation techniques that recognize the unique characteristics of insured losses.

Acknowledgements

The authors would like to thank the Insurance Services Office for their generosity in supplying the industry data used in this analysis, the ratemaking call paper committee for their guidance and our families for their understanding and support throughout the process of drafting and editing this paper.

Index to Exhibits

Nevada Bodily Injury
Insurance Industry Loss Data

YY/Q	Qtrly Paid Frequency	Qtrly Paid Severity
94/1	2.018	8,836.39
94/2	2.042	8,634.60
94/3	2.100	9,021.44
94/4	2.186	8,310.44
95/1	2.108	8,000.58
95/2	2.140	8,040.02
95/3	1.967	8,786.99
95/4	2.064	9,415.44
96/1	1.954	7,993.37
96/2	1.842	9,213.77
96/3	1.751	9,124.03
96/4	1.757	9,084.54
97/1	1.739	8,371.74
97/2	1.861	9,572.92
97/3	1.837	8,560.24
97/4	1.831	9,103.45
98/1	1.770	11,106.61
98/2	1.999	9,743.20
98/3	1.778	8,651.21
98/4	1.749	9,552.60
99/1	1.799	9,594.95
99/2	1.830	9,205.35
99/3	1.755	9,799.76

YY/Q	Four Qtr Ending Paid Frequency	Four Qtr Ending Paid Severity
94/1		
94/2		
94/3		
94/4	2.087	8,694.21
95/1	2.110	8,486.04
95/2	2.134	8,338.22
95/3	2.100	8,277.70
95/4	2.070	8,557.51
96/1	2.031	8,560.77
96/2	1.956	8,855.56
96/3	1.901	8,935.94
96/4	1.825	8,840.39
97/1	1.772	8,950.26
97/2	1.778	9,049.11
97/3	1.799	8,904.46
97/4	1.817	8,912.80
98/1	1.824	9,574.94
98/2	1.859	9,621.17
98/3	1.844	9,642.72
98/4	1.823	9,753.45
99/1	1.830	9,396.03
99/2	1.789	9,252.20
99/3	1.783	9,535.72

New York Collision
Insurance Industry Loss Data

YY/Q	Qtrly Paid Frequency	Qtrly Paid Severity
94/1	10.085	1,969.88
94/2	7.458	1,753.67
94/3	7.359	1,946.69
94/4	7.586	2,073.42
95/1	7.951	2,150.86
95/2	6.858	2,022.18
95/3	7.067	2,106.83
95/4	7.692	2,214.01
96/1	9.326	2,230.18
96/2	6.993	2,037.11
96/3	6.948	2,113.95
96/4	7.575	2,275.75
97/1	7.792	2,460.61
97/2	6.860	2,185.90
97/3	7.023	2,226.98
97/4	7.235	2,301.27
98/1	7.423	2,349.35
98/2	6.835	2,112.72
98/3	6.889	2,225.14
98/4	6.982	2,268.54
99/1	8.103	2,392.75
99/2	6.821	2,196.11
99/3	7.000	2,291.83

YY/Q	Four Qtr Ending Paid Frequency	Four Qtr Ending Paid Severity
94/1		
94/2		
94/3		
94/4	8.117	1,939.10
95/1	7.588	1,984.33
95/2	7.437	2,050.69
95/3	7.363	2,090.35
95/4	7.391	2,126.97
96/1	7.738	2,152.11
96/2	7.768	2,154.42
96/3	7.734	2,155.96
96/4	7.704	2,171.35
97/1	7.328	2,230.23
97/2	7.292	2,265.26
97/3	7.309	2,291.61
97/4	7.225	2,298.13
98/1	7.137	2,268.21
98/2	7.128	2,249.98
98/3	7.094	2,249.49
98/4	7.031	2,241.10
99/1	7.200	2,255.85
99/2	7.198	2,275.84
99/3	7.226	2,291.96

Oregon Homeowner
Insurance Industry Loss Data

YY/Q	Qtrly Paid Frequency	Qtrly Paid Severity
94/1	6.167	2,365.18
94/2	5.778	2,228.65
94/3	6.194	2,224.27
94/4	7.319	2,227.28
95/1	7.573	2,477.43
95/2	6.665	2,436.19
95/3	8.076	2,700.68
95/4	8.613	2,209.83
96/1	24.861	1,973.35
96/2	8.456	2,620.94
96/3	7.006	2,832.13
96/4	6.555	3,070.97
97/1	9.303	2,353.67
97/2	6.053	2,535.58
97/3	5.906	2,747.17
97/4	5.778	2,556.34
98/1	7.300	2,689.08
98/2	5.301	2,569.03
98/3	5.592	3,034.36
98/4	5.986	2,730.10
99/1	8.539	3,126.60
99/2	5.463	3,313.96
99/3	4.965	3,625.18

YY/Q	Four Qtr Ending Paid Frequency	Four Qtr Ending Paid Severity
94/1		
94/2		
94/3		
94/4	6.366	2,259.55
95/1	6.715	2,297.02
95/2	6.935	2,344.92
95/3	7.409	2,468.28
95/4	7.734	2,452.25
96/1	12.069	2,200.97
96/2	12.493	2,241.76
96/3	12.196	2,252.52
96/4	11.656	2,376.12
97/1	7.827	2,683.58
97/2	7.222	2,670.32
97/3	6.942	2,646.38
97/4	6.744	2,525.25
98/1	6.258	2,635.14
98/2	6.066	2,645.05
98/3	5.984	2,713.38
98/4	6.035	2,754.39
99/1	6.353	2,897.10
99/2	6.384	3,054.73
99/3	6.220	3,175.16

Oregon Homeowners
Exponential Regression with Indicator Variables on Quarterly Frequency

Observation	YY/Q	Qtrly Freq.	X_i	D2	D3	D4	Ln(Freq)	ln(Fitted Freq	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq.
1	94/4	6.17	0.00	0	0	1	1.820	2.307	-0.4875	0.24	-	10.05
2	95/1	5.78	0.25	0	0	0	1.754	1.903	-0.1485	0.02	0.11	6.71
3	95/2	6.19	0.50	1	0	0	1.823	1.924	-0.1007	0.01	0.00	6.85
4	95/3	7.32	0.75	0	1	0	1.991	1.966	0.0249	0.00	0.02	7.14
5	95/4	7.57	1.00	0	0	1	2.024	2.281	-0.2569	0.07	0.08	9.79
6	96/1	6.66	1.25	0	0	0	1.896	1.877	0.0193	0.00	0.08	6.53
7	96/2	8.08	1.50	1	0	0	2.089	1.898	0.1918	0.04	0.03	6.67
8	96/3	8.61	1.75	0	1	0	2.153	1.940	0.2133	0.05	0.00	6.96
9	96/4	24.86	2.00	0	0	1	3.213	2.255	0.9583	0.92	0.55	9.54
10	97/1	8.46	2.25	0	0	0	2.135	1.851	0.2846	0.08	0.45	6.36
11	97/2	7.01	2.50	1	0	0	1.947	1.871	0.0759	0.01	0.04	6.50
12	97/3	6.55	2.75	0	1	0	1.879	1.913	-0.0340	0.00	0.01	6.78
13	97/4	9.30	3.00	0	0	1	2.230	2.229	0.0011	0.00	0.00	9.29
14	98/1	6.05	3.25	0	0	0	1.800	1.825	-0.0246	0.00	0.00	6.20
15	98/2	5.91	3.50	1	0	0	1.777	1.845	-0.0687	0.00	0.00	6.33
16	98/3	5.78	3.75	0	1	0	1.754	1.887	-0.1330	0.02	0.00	6.60
17	98/4	7.30	4.00	0	0	1	1.988	2.203	-0.2150	0.05	0.01	9.05
18	99/1	5.30	4.25	0	0	0	1.668	1.799	-0.1308	0.02	0.01	6.04
19	99/2	5.59	4.50	1	0	0	1.721	1.819	-0.0983	0.01	0.00	6.17
20	99/3	5.99	4.75	0	1	0	1.790	1.861	-0.0712	0.01	0.00	6.43

Sum	1.53	1.41
D	**0.92**	
Number of X	4.00	
Observations	20.00	

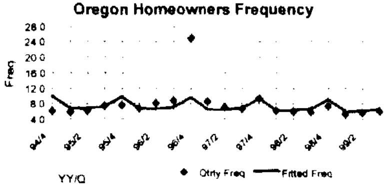
d_U at .05	1.83
d_L at .05	0.90

Test is Inconclusive

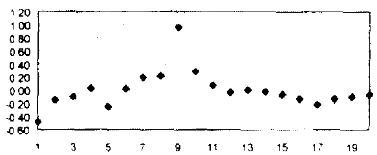
d_U at .01	1.57
d_L at .01	0.68

Test is Inconclusive

Oregon Homeowners Frequency



Residual Plot



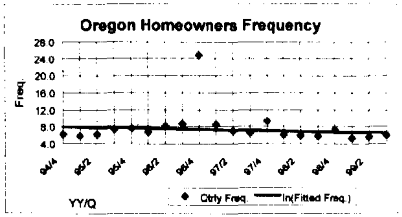
Regression Output

Trend	-2.58%
R^2	0.27
Obs.	20

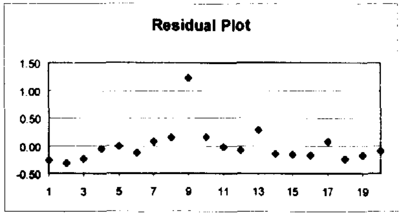
Seasonality Factors	
1st Qtr	1.000
2nd Qtr	1.028
3rd Qtr	1.079
4th Qtr	1.488

Oregon Homeowners
Exponential Regression on Quarterly Frequency

Observation	YY/Q	Qtrly Freq.	Xi	Ln(Freq)	ln(Fitted Freq)	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq.
1	94/4	6.17	0.00	1.820	2.068	-0.2485	0.06	-	7.91
2	95/1	5.78	0.25	1.754	2.058	-0.3037	0.09	0.00	7.83
3	95/2	6.19	0.50	1.823	2.048	-0.2251	0.05	0.01	7.75
4	95/3	7.32	0.75	1.991	2.038	-0.0474	0.00	0.03	7.68
5	95/4	7.57	1.00	2.024	2.028	-0.0038	0.00	0.00	7.60
6	96/1	6.66	1.25	1.896	2.018	-0.1218	0.01	0.01	7.52
7	96/2	8.08	1.50	2.089	2.008	0.0815	0.01	0.04	7.45
8	96/3	8.61	1.75	2.153	1.998	0.1551	0.02	0.01	7.37
9	96/4	24.86	2.00	3.213	1.988	1.2255	1.50	1.15	7.30
10	97/1	8.46	2.25	2.135	1.978	0.1577	0.02	1.14	7.23
11	97/2	7.01	2.50	1.947	1.968	-0.0203	0.00	0.03	7.15
12	97/3	6.55	2.75	1.879	1.958	-0.0781	0.01	0.00	7.08
13	97/4	9.30	3.00	2.230	1.948	0.2825	0.08	0.13	7.01
14	98/1	6.05	3.25	1.800	1.937	-0.1374	0.02	0.18	6.94
15	98/2	5.91	3.50	1.777	1.927	-0.1508	0.02	0.00	6.87
16	98/3	5.78	3.75	1.754	1.917	-0.1630	0.03	0.00	6.80
17	98/4	7.30	4.00	1.988	1.907	0.0805	0.01	0.06	6.74
18	99/1	5.30	4.25	1.668	1.897	-0.2296	0.05	0.10	6.67
19	99/2	5.59	4.50	1.721	1.887	-0.1663	0.03	0.00	6.60
20	99/3	5.99	4.75	1.790	1.877	-0.0871	0.01	0.01	6.54



Sum	2.03	2.90
D	**1.43**	
Number of X	1.00	
Observations	20.00	
d_U at .05	1.41	
d_L at .05	1.20	
Uncorrelated		
d_U at .01	1.15	
d_L at .01	0.95	
Uncorrelated		



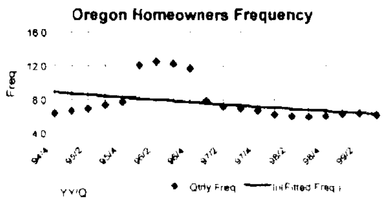
Regression Output

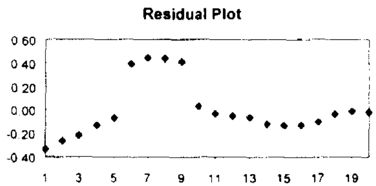
Trend	-3.94%
R^2	0.03
Obs.	20

Oregon Homeowners
Exponential Regression on 4QE Frequency

Observation	YY/Q	4QE Freq.	Xi	Ln(Freq)	ln(Fitted Freq)	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq.
1	94/4	6.37	0.00	1.852	2.187	-0.3355	0.11	-	8.91
2	95/1	6.72	0.25	1.905	2.169	-0.2641	0.07	0.01	8.75
3	95/2	6.93	0.50	1.936	2.151	-0.2155	0.05	0.00	8.60
4	95/3	7.41	0.75	2.003	2.133	-0.1306	0.02	0.01	8.44
5	95/4	7.73	1.00	2.045	2.116	-0.0705	0.00	0.00	8.29
6	96/1	12.07	1.25	2.491	2.098	0.3930	0.15	0.21	8.15
7	96/2	12.49	1.50	2.525	2.080	0.4450	0.20	0.00	8.00
8	96/3	12.20	1.75	2.501	2.062	0.4394	0.19	0.00	7.86
9	96/4	11.66	2.00	2.456	2.044	0.4120	0.17	0.00	7.72
10	97/1	7.83	2.25	2.058	2.026	0.0316	0.00	0.14	7.59
11	97/2	7.22	2.50	1.977	2.008	-0.0316	0.00	0.00	7.45
12	97/3	6.94	2.75	1.937	1.991	-0.0533	0.00	0.00	7.32
13	97/4	6.74	3.00	1.908	1.973	-0.0647	0.00	0.00	7.19
14	98/1	6.26	3.25	1.834	1.955	-0.1207	0.01	0.00	7.06
15	98/2	6.07	3.50	1.803	1.937	-0.1337	0.02	0.00	6.94
16	98/3	5.98	3.75	1.788	1.919	-0.1307	0.02	0.00	6.82
17	98/4	6.04	4.00	1.798	1.901	-0.1029	0.01	0.00	6.69
18	99/1	6.35	4.25	1.848	1.883	-0.0350	0.00	0.00	6.58
19	99/2	6.38	4.50	1.853	1.866	-0.0124	0.00	0.00	6.46
20	99/3	6.22	4.75	1.828	1.848	-0.0199	0.00	0.00	6.35
						Sum	1.04	0.40	

D	**0.38**
Number of X	1.00
Observations	20.00
d_U at .05	1.41
d_L at .05	1.20
First Order Auto-Correlated	
d_U at .01	1.15
d_L at .01	0.95
First Order Auto-Correlated	





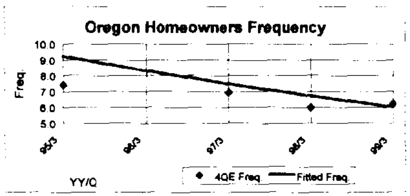
Regression Output

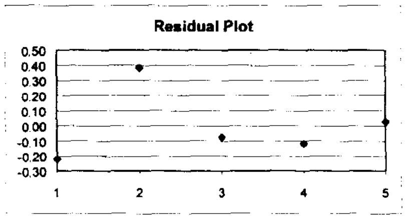
Trend	-6.89%
R^2	0.17
Obs.	20

Oregon Homeowners
Exponential Regression 4QE Frequency, Annual Observations

Observation	YY/Q	4QE Freq	Xi	Ln(Freq)	ln(Fitted Freq	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq
1	95/3	7.41	0.00	2.003	2.224	-0.2213	0.05	-	9.25
2	96/3	12.20	1.00	2.501	2.118	0.3836	0.15	0.37	8.31
3	97/3	6.94	2.00	1.937	2.012	-0.0742	0.01	0.21	7.47
4	98/3	5.98	3.00	1.788	1.905	-0.1168	0.01	0.00	6.72
5	99/3	6.22	4.00	1.828	1.799	0.0288	0.00	0.02	6.04

Sum	0.22	0.60
D	2.77	
Number of X	1.00	
Observations	5.00	
dU at .05	na	
dL at .05	na	
dU at .01	na	
dL at .01	na	

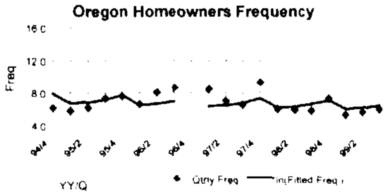




Oregon Homeowners

Manually Adjusted Exponential Regression with Indicator Variables on Quarterly Frequency

										Durbin-Watson		
Observation	YY/Q	Qtrly Freq	Xi	D2	D3	D4	Ln(Freq)	Fitted Freq	Residuals	e_t^2	$(e_t-e_{t-1})^2$	Fitted Freq
1	94/4	6.17	0.00	0	0	1	1.820	2.068	-0.2479	0.06	-	7.91
2	95/1	5.78	0.25	0	0	0	1.754	1.903	-0.1485	0.02	0.01	6.71
3	95/2	6.19	0.50	1	0	0	1.823	1.924	-0.1007	0.01	0.00	6.85
4	95/3	7.32	0.75	0	1	0	1.991	1.966	0.0249	0.00	0.02	7.14
5	95/4	7.57	1.00	0	0	1	2.024	2.042	-0.0173	0.00	0.00	7.70
6	96/1	6.66	1.25	0	0	0	1.896	1.877	0.0193	0.00	0.00	6.53
7	96/2	8.08	1.50	1	0	0	2.089	1.898	0.1918	0.04	0.03	6.67
8	96/3	8.61	1.75	0	1	0	2.153	1.940	0.2133	0.05	0.00	6.96
10	97/1	8.46	2.25	0	0	0	2.135	1.851	0.2846	0.08	0.01	0.00
11	97/2	7.01	2.50	1	0	0	1.947	1.871	0.0759	0.01	0.04	6.36
12	97/3	6.55	2.75	0	1	0	1.879	1.913	-0.0340	0.00	0.01	6.50
13	97/4	9.30	3.00	0	0	1	2.230	1.989	0.2407	0.06	0.08	6.78
14	98/1	6.05	3.25	0	0	0	1.800	1.825	-0.0246	0.00	0.07	7.31
15	98/2	5.91	3.50	1	0	0	1.777	1.845	-0.0687	0.00	0.00	6.20
16	98/3	5.78	3.75	0	1	0	1.754	1.887	-0.1330	0.02	0.00	6.33
17	98/4	7.30	4.00	0	0	1	1.988	1.963	0.0246	0.00	0.02	6.60
18	99/1	5.30	4.25	0	0	0	1.668	1.799	-0.1308	0.02	0.02	7.12
19	99/2	5.59	4.50	1	0	0	1.721	1.819	-0.0983	0.01	0.00	6.04
20	99/3	5.99	4.75	0	1	0	1.790	1.861	-0.0712	0.01	0.00	6.17
												6.43
									Sum	0.38	0.32	
									D	**0.86**		
									Number of X	4.00		
									Observations	19.00		

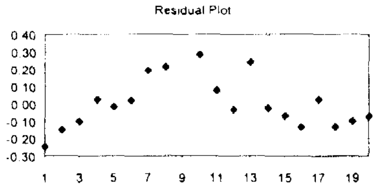


d_U at .05	1.85
d_L at .05	0.86

First Order Auto-Correlated

d_U at .01	1.58
d_L at .01	0.65

Test is Inconclusive



Regression Output

Trend	-2.58%
R^2	0.20
Obs	19

Seasonality Factors	
1st Qtr	1.000
2nd Qtr	1.028
3rd Qtr	1.079
4th Qtr	1.171

Ratemaking for Excess Workers Compensation

Owen M. Gleeson, FCAS, MAAA

Ratemaking for Workers Compensation

By Owen M. Gleeson, FCAS, MAAA

Abstract

The market for Excess Workers Compensation in the United States has grown rapidly over the last two decades. These are estimates that the annual premium volume in the excess $500,000 attachment segment of this market is now in excess of $1 billion. This paper presents a method of estimating rates for this type of coverage. The method generates loss distribution of the total cost of individual large claims. Medical costs are estimated form data samples. Indemnity costs, however, are for the most part estimated from the benefits mandated in the Workers Compensation statutes.

I. Introduction

A. General Remarks

The market for property/casualty insurance in the United States has evolved rapidly in the past 15 years. In particular, the alternative market for Workers compensation insurance has shown explosive growth. Many of the entities that incur workers compensation costs are now self-insured on the lower cost layer, e.g. the first $100,000 per claim. These self-insured firms or groups still purchase insurance protection above retentions that are $100,000 higher. The market for this type of coverage is now very large and in premium dollar terms easily exceeds $1 billion. Another measure of the size of the market is that the Self-Insurance Institute of America has over one thousand corporate members.

The task of estimating rates for this type of business is made difficult by several of the characteristics of large workers compensation claims. The first is that large workers compensation claims are infrequent and thus the amount of data available for ratemaking is severely limited. A second characteristic is that large workers compensation claims develop very slowly with the result that the ultimate cost of an individual claim, particularly those involving medical may not be known for many years. Another aspect of these claims is that there are distinct components of the loss:medical and indemnity. The view adopted here is that the medical costs and the indemnity costs follow separate and distinct distributions. As a result the distribution of the variable which is the sum of these costs is quite complex. It is thus very difficult to model the underlying distribution of these costs by using a sample of incurred losses.

Currently there is no pricing mechanism in the United States for this class of business that provides comfort to the users and is widely accessible. The objective of this paper is to provide a solution to the problem of pricing this line of business which will be seen as generally satisfactory. There are of course no claims implied that what is presented in the following is the only solution or the solution that is "best" in some sense. In addition, this paper will not explore the issue of risk loading or required profit. Rather the paper will focus on the sufficiently difficult task of estimating the pure loss cost.

B. Types of Claims

The focus of this paper is excess workers compensation costs. It follows that only those types of claims whose cost might exceed a given limit e.g. $100,000 would be of interest. Workers Compensation claims are often classified into six types: Medical Only, Temporary Total, Minor Permanent Partial, Major Permanent Partial, Permanent Total and Fatal. It's assumed for the purposes of this paper that no claim falling into one of the first three classifications will be large enough to pierce the limits of interest. Therefore only the remaining three types of claims will be analyzed.

At this point a discussion of the characteristics of each of the three types of claims will be presented. It is hoped that this will provide motivation for the methods and tactics used in producing the cost estimates. Each of the types of claims to be discussed, i.e. Fatal, Permanent Total, and Major Permanent Partial have a medical component of the total claim cost and an indemnity component of the cost. These will be discussed separately.

1. Fatal
   a. Indemnity Benefits

The statutory specification of the indemnity benefits associated with fatal claims can be quite complex. In highly simplified terms the parameters specifying the benefits might be described as (1) period of benefits (2) basic percentage of wage and (3) degree of dependency. For example, the period of benefits could be lifetime. However the period of benefits could be limited by attained age, say age 65, or limited by amount (the maximum amount of fatal benefits in Florida is $100,000). The basic percentage of wage is usually expressed in terms such as "66 2/3 percent of the fatally injured individuals average weekly wage." (Many workers in the United States do not receive the same amount of compensation every week. As a result, it is necessary to determine the amount that should be deemed the average weekly wage in the event of injury. Each state has developed a complex set of rules to decide this question. This subject will not be explored here.) The degree of dependency in a fatal case is determined generally by familial status e.g. spouse, spouse and dependent children, dependent parents or siblings, etc.

The specifications vary from one state to another. Thus the first step in dealing with the costs of fatal claims is to analyze the laws of the state for which rates are being estimated. Another step in the process is to decide on the simplifying assumptions that need to be made in order to make the calculations tractable.

An example of the detail that should be considered in analyzing the excess workers compensation costs for a given state is the mandates of the State of Pennsylvania. This is to be found in Appendix A.

   b. Medical Costs

It would be reasonable to think that there are probably little or no medical costs associated with a Fatal claim. However, the data sets that the author and his associates have reviewed have virtually all presented some fatal claims with related medical costs. For the majority of fatal claims the medical cost is found to be zero. However, there are medical costs associated with the other fatal claims and these seem to fall into the following categories: small, medium and very large. We speculate that the small costs are ambulance and emergency rooms fees for individuals who survive a matter of hours. The medium costs may be associated with claims where the injured party survived for a matter of days and then expired.

The very large costs were likely the result of heroic and extensive efforts to treat a very seriously injured person with the result that life was sustained for a year or two followed by the expiration of the injured person. This last group averages over $1,000,000, but seems extremely rare.

The above view has been developed by examining claim files, discussions with claims adjusters and from conversations with others personally familiar with the details of high cost workers compensation claims

2. Permanent Total

a. Indemnity

As in the case with fatal claims, the statutory specifications of indemnity benefits due an impaired party can be fairly complex. The general parameters are 1) the period of benefits 2) limitations and/or offsets and 3) basic percentage of wage. The period of benefits for permanent total claims in most states is lifetime. Many states mandate payment of full benefits to injured individuals as long as they survive. However in other states there are limitations or offsets most of which are associated in one way or another with Social Security. For example, some states mandate payments only until eligibility for Social Security. On the other hand some states require that the basic benefits be offset by benefits obtainable under the Disability provisions of Social Security. The offsets vary widely from state to state and can have significant impact on the cost of permanent total claims. Finally there is the question of the basic percentage. This is usually expressed as something like 66 2/3 percent of wages. However the percent is different from one state to another and may be expressed as a percent of spendable income.

Again the law of the state under consideration must be analyzed carefully. Also as is the case with fatal claims, it may be necessary to make some simplifying assumptions.

b. Medical

Many Permanent Total claims are characterized by extremely large medical costs. Not only are the costs large but the costs seem to develop upwards throughout the life of the claim which may be on the order of several decades. Unfortunately, most data collecting agencies do not follow the development on individual claims for a sufficiently long time. This is not to be construed as criticism but rather recognizes the fact that the development in PT claims while perhaps very large for an individual claim may not contribute a significant amount of development to the overall workers compensation total loss cost. As an example, if the developed medical cost on PT's through say 10 years is 4% of the total loss cost dollar and the remaining development is 50% (probably too negative a view) then the overall pure premium might be underestimated by 2%.

However, the interest here is not in aggregates but in the size of individual claims. The data used by the author is drawn from a number of private well-maintained databases of individual workers compensation claims. In each of these, there are claims from many accident years. The open claims are developed individually. The method will be addressed in a later section. Both closed and open are then trended to the experience period. Since Permanent Total claims are rather rare it seems virtually impossible to generate a data set that can be used to provide an empirical size of loss distribution that can be used without resorting to some smoothing. Thus, some smoothing (graduation) must be introduced before the "tail" of the distribution can be used for pricing.

3. Major Permanent Partial

It's customary in Workers Compensation data preparation to rely on "C" values to distinguish between Major Permanent Partial and Minor Permanent Partial. The problem with using this definition is that the c-values vary by state and by accident year.

The approach used here was to obtain data by state on claims designated Major Permanent Partial and to examine the characteristics of the data. This was supplemented by information drawn from Workers Compensation Loss Cost filings from New York and Pennsylvania which contain considerable detail State Workers Compensation laws were also consulted with respect to benefits provided for permanent partial.

Evaluation of this body of information led to conclusions with respect to the medical distribution and the indemnity distribution. The expected value and the range of the distribution as well as some general characteristics are discussed in the following.

a. Indemnity

The indemnity associated with a Permanent Partial claim generally depends on the type of injury. Examples of the type of injury are "Loss of a hand", "loss of an arm", "Loss of a foot", and so forth. An example of the compensation is "Loss of a hand –335 weeks". The amount of compensation is usually a percent of wage, e.g. 66 2/3 percent. As shown in Appendix A, state workers compensation law list many specific types of injury each of which entitles the injured party to a particular set of benefits.

The large number of categories alone would make modelling of the costs difficult even if there were good data on the frequency of type of injury. However this is not the case. In addition, analysis indicates that Permanent Partial claims do not contribute significantly to the overall excess costs. This is due to the fact that review of an extensive amount of data shows that, while the Permanent Partial claims are serious with a large average value, the frequency of claims in excess of say $500,000 is low and that there are also no truly catastrophic claims.

Given the above it was decided to resort to analysis of sample data to estimate the distribution of indemnity of Major Permanent Partial claims.

b. Medical

Indemnity costs on Major Permanent Partial are relatively well constrained by the limitations resulting from statutorily defined benefits. However, injuries resulting in Permanent Partial disability can result in a large range of incurred medical costs. In some cases, such as loss of a hand, the injury maybe satisfactorily treated rapidly and at a low medical cost. On the other hand there are catastrophic injuries such as severe burns or injuries to the spinal column where the injured party will require significant medical treatment but will eventually be able to return to work. At this point it might be observed that there are some individuals who find that their quality of life is enhanced if they are able to resume some sort of gainful employment no matter how serious the injury. Thus, these individuals cannot be considered to be permanently and totally disabled.

C. Discussion of Lack of Data

It's probably worthwhile at this point to recall that the objective is to determine rates for excess workers compensation coverages. Thus by far the largest number of claims incurred under Workers Compensation coverages are, by definition, of no interest. For example, consider the following data extracted from a Pennsylvania Compensation Rating Bureau Loss Cost filing.

Table 1
Ultimate Number of Injuries

Period	Fatal	Permanent Total	Major Permanent Partial	Minor Permanent Partial	Temporary Total
1991	150	207	2480	3411	39,571
1992	167	205	2449	3375	39,124
1993	132	201	2393	3304	38,154
1994	163	203	2394	3308	38,093
1995	110	211	2449	3406	39,004
	722	1,027	12,165	16,804	193,946

From the point of view of credibility standards, it can be seen that there are insufficient claims of the type of interest for rate making purposes even if the claims were restricted to basic limits as found in other lines of business. Of course as previously mentioned the size of some of the claims encountered in Statutory Workers Comp range up to $20 million. While it would be interesting to determine the number of claims necessary for full credibility on claims of this size the knowledge gained is probably not worth the effort. However, we suspect that it is well in excess of all the claims of the size under consideration that are incurred in the United States in the span of a decade. Thus the answer is irrelevant since the number required exceeds the number available. Therefore it is necessary to develop an approach that circumvents this lack of data.

Excess Workers comp rates are needed by state since the statutory benefits vary by state with respect to the indemnity portion of the claim. This compounds the data availability problem in that a smaller number of claims are available in a given jurisdiction. Also whereas relatively large states like Pennsylvania and Texas which have respective populations of approximately 12 million and 17 million might have enough claims to provide basis for a reasobably accurate estimate, the problem of constructing rates for states like Iowa and Oregon with populations of approximately 3 million each remains.

Another issue that surfaced in the process of the construction of the rates is that the indemnity portions of the serious workers compensation claim develops much differently from the medical portion of the claim. The data in the table below has been generated by using data drawn from a recent Pennsylvania Loss Cost filing to demonstrate that the indemnity costs develop much more rapidly that medical costs. This stands to reason. Consider a typical Permanent Total claim. Within a matter of five to ten years it should be certain that the claimant is entitled to Permanent Total benefits. At this point, the cost of the indemnity portion of the claim has been precisely determined. However the medical costs are a function of how well the claimant responds to treatment, indicated alternative treatment paths that emerge, new developments in medical care and so forth.

	Required Reserve/ Current Reserve	
Period	Medical	Indemnity
12 months	2.8180	3.6879
24 months	2.6563	2.3059
36 months	2.7917	1.6556
48 months	2.8603	1.3140
60 months	3.1292	1.1207
72 months	3.2063	1.0000 *

*approximate

The above suggests that applying a single development factor to the total of indemnity and medical will likely produce less satisfactory results than the process of applying development factors separately if possible or avoiding the use of development factors if feasible.

Another aspect of the data problem is the question of combining data from different states. Because the indemnity benefits (which account for about 50% of Major Permanent Partial and 2/3 of Permanent Total costs) vary so significantly from one state to another as a result of offsets, limitations, etc not to mention escalation it was decided that the approach that would produce the most accurate results would be to estimate the indemnity costs by state if at all possible.

On the other hand medical costs are not statutorily determined. While costs of some of the more minor aspects of medical care such as bandages, splints, emergency room costs probably display regional variations, the larger dollar costs such as treatment at national burn care units or spinal treatment centers demonstrate more homogeneity than indemnity. In addition the treatment proposed for estimating state indemnity costs has no analogue for medical cost.

The above characteristics of serious workers compensation claims: low frequency, high severity, different types of development for component costs and lack of comparability of cost from state to state led to the solution proposed on the next section.

II. General Approach to Solution of Estimating Excess Workers Compensation Costs

A. Outline of Basic Solution

The basic solution to modeling the distribution of costs of large claims consists of two steps. The first step was to create a distribution of costs for each type of serious claim: Fatal, Permanent Total and Major Permanent partial. This step required the creation of separate of distributions for indemnity and medical. These distributions were then used to create a joint distribution for each of the type of claims. Excess cost factors are then generated for each type of claim.

The second step was to determine the portion of the pure premium that is Fatal, Permanent Total or Major Permanent Partial and then weight the excess factors of the individual components.

The statement of the solution is fairly simple. However, the physical execution of it is not. For example, given the above, the number of cost outcomes or cells for Permanent Total Costs is numbered in the millions using an approximating method of calculating the costs. Essentially what is determined is the frequency function CPT (m,w,a,l) where m is medical cost, w is wage, a is age at time of injury and l is the number of years lived after the injury. The distribution of the costs of fatal claims $CF(m,w,a,l)$ is calculated in a similar manner. The cost distribution for Major Permanent Partial is obtained in a slightly different manner. One component is the medical cost. The other is the indemnity. However the awards are not so life or age dependent since there are certain lump sums statutorily provided for regardless of age or wage. Thus for this type of injury the distribution of indemnity is determined from a statistical sample. The compound distribution of costs is denoted $CMPP(m, I)$.

For a given retention, R, the excess costs as a percentage of total costs are obtained by type of injury for a given state. These percents are then weighted by the percent of the pure premium ascribable to that type of injury. For example, suppose the retention for State G is 500,000. Further suppose that 58.8% of total PT costs are excess 500,000; 2.48% of total Fatal costs are excess 500,000 and 3.36% of Major Permanent Partial are excess 500,000. Also suppose that 12.2% of the pure premium (loss cost only) is the cost of PT's, 3.1% is the Fatal cost and 63.3% is the Major PP cost with 21.2% of loss costs attributable to other types of injuries.

Then the excess factor for 500,000
is (58.8%)(12.2%)+(2.48%)(3.1%)+(3.36%)(63.3%) = 9.38%

The problem to be solved, the difficulties, motivation and methodology have been outlined above. What follows are some examples that are designed to assist in the understanding of the methodology.

B. Examples

1. Example #1

In this example it is assumed that there are three types of claims which account collectively for all the incurred loss. The goal is to determine that excess costs for an attachment point of $500,000. Each type of claim is comprised of two components. The components are considered to be independent. The distribution of the components of each type of claim are given in the tables below.

Claim Type 1

Component A(1)		Component B(1)	
Amount	Prob.	Amount	Prob.
100,000	6.0%	150,000	8.0%
150,000	8.0%	225000	9.0%
200,000	9.0%	300000	12.0%
250,000	10.0%	375,000	14.0%
300,000	12.0%	450,000	18.0%
350,000	14.0%	525,000	16.0%
400,000	16.0%	600,000	14.0%
450,000	11.0%	1000000	6.0%
500,000	9.0%	1500000	2.0%
550,000	5.0%	2,000,000	1.0%

Claim Type 2

Component A(2)		Component B(2)	
Amount	Prob.	Amount	Prob.
25,000	2.0%	50,000	4.0%
75,000	3.0%	100,000	6.0%
125,000	5.0%	150,000	10.0%
175,000	15.0%	200,000	12.0%
225,000	25.0%	250,000	18.0%
275,000	25.0%	300,000	18.0%
325,000	15.0%	350,000	12.0%
375,000	5.0%	400,000	10.0%
425,000	3.0%	450,000	6.0%
475,000	2.0%	500,000	4.0%

Claim Type 3

Component A(3)		Component B(3)	
Amount	Prob.	Amount	Prob.
0	85.0%	200,000	8.0%
50,000	10.0%	250,000	9.0%
100,000	4.0%	300,000	10.0%
500,000	1.0%	350,000	11.0%
		400,000	12.0%
		450,000	12.0%
		500,000	11.0%
		550,000	10.0%
		600,000	9.0%
		650,000	8.0%

If a joint distribution is created for each type of claim and the excess of 500,000 percent is calculated for each, the excess percent is as shown in the following table.

Excess Cost

Claim Type	Prcnt.
#1	39.50%
#2	13.40%
#3	7.70%

Further assume that percent of the pure premium is known to be distributed as follows

Distribution of Loss Cost

Claim Type	Prcnt.
#1	5.2%
#2	71.3%
#3	23.5%

Then the percent of the cost excess 500,000 is calculated as
(39.5%)(5.2%)+(13.4%)(71.3%)+(7.7%)(23.5%) = 16.55%

2. Example #2

This example illustrates some of the calculations involved in estimating the distribution of costs for Fatal claims. In order to estimate the distribution of indemnity costs for a fatal claim a number of parameters need to be specified. These are as follows:

a. Wage Distribution

Ratio AWW to SAWW*	Percent Workers Earning AWW
0.30	5.0%
0.60	30.0%
1.00	40.0%
1.35	10.0%
1.50	15.0%

*AWW = Average Weekly Wage, SAWW = State Average Weekly Wage

b. State Average Weekly Wage

SAWW = $600

c. Distribution of Ages at time of death

Age	Percent of Workers at Age
20	20.0%
30	20.0%
40	20.0%
50	20.0%
60	20.0%

d. Benefit Assumptions

Surviving spouse receives 66 2/3% of wage at time of death.

Maximum = 100% SAWW
Minimum = 20% SAWW

e. Life Table

US Life Table – 1980
See Appendix B.

f. Distribution of Indemnity Costs

Given data in a., b., d., and e. above the distribution of indemnity costs for fatalities suffered by individuals aged 40 is as given in the following table.

Distribution of
Indemnity Costs
at Age 40

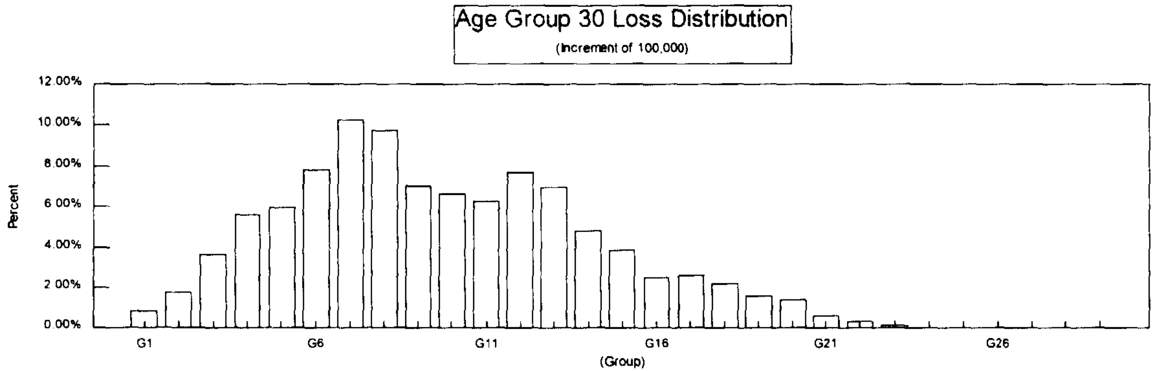
Group	Probability	Amount	Group	Probability	Amount
G1	1.78%	60,545	G13	3.65%	1,245,619
G2	3.88%	158,419	G14	2.54%	1,343,204
G3	6.76%	256,890	G15	2.80%	1,447,386
G4	7.94%	354,637	G16	1.59%	1,553,684
G5	9.89%	458,065	G17	1.15%	1,646,886
G6	11.19%	554,423	G18	0.64%	1,743,391
G7	10.20%	647,978	G19	0.30%	1,836,176
G8	7.89%	748,037	G20	0.15%	1,938,251
G9	7.96%	854,019	G21	0.03%	2,051,441
G10	8.30%	955,471	G22	0.01%	2,144,589
G11	6.21%	1,051,717	G23	0.003%	2,215,200
G12	5.10%	1,144,313			

The figures in the above table were obtained by first calculating the costs for each individual cell. For example, suppose a fatally injured worker was earning $810 per week. Then the surviving spouse's weekly benefits would be (662/3%)($810) = $540 or an annual amount of $28,080. Also assume that the spouse receives benefits for exactly twenty years and then dies. The amount received is (20)($28,080) = $561,600 and the probability of this event is (10%)((84,789 - 83,726)/94,926) = .112% (see wage distribution and Appendix B). The outcomes were then grouped into intervals of $100,000. The outcome of the above described event would fall into group G6.

A graph of the distribution of indemnity costs for a person age 40 is shown in Figure #1. This is followed by a graph of the distribution of costs for a person age 30 in Figure #2.

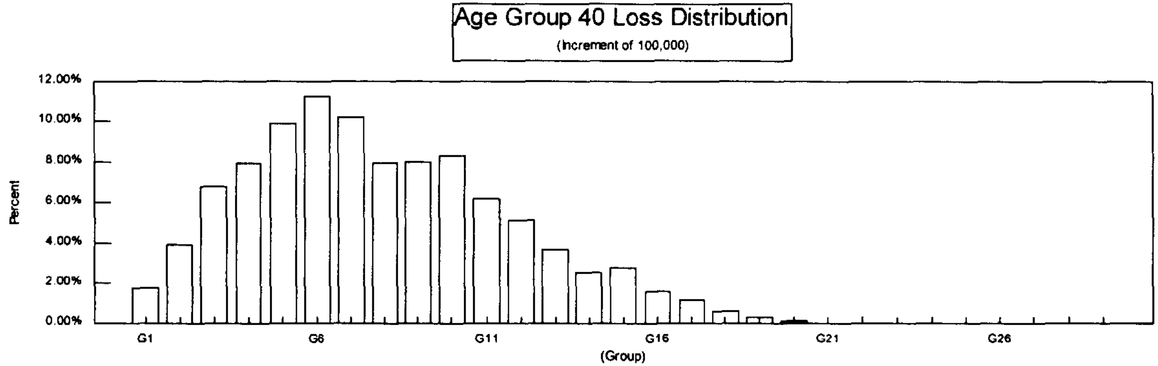
A few things should be noted about the two graphs. One is that the distribution of costs in the age 30 graph is somewhat to the right of the age 40 distribution. This would be intuitively expected since the individuals age 30 at time of death would provide about an additional 10 years of benefits to their survivors.

Figure #1



Age Group 30 Loss Distribution
(Increment of 100,000)
Percent
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
G1
G6
G11
G16
G21
G26
(Group)

Figure #2



Age Group 40 Loss Distribution
(Increment of 100,000)
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
Percent
G1
G6
G11
G16
G21
G26
(Group)

It's also interesting to note that both of the distributions are somewhat "lumpy". The distributions have been created from a life table which is fairly smooth, and the combination of a wage distribution and certain benefits assumptions.

It seems that the fact that the wage distribution shows an uneven distribution of wages and the statutory benefits display certain maximums and minimums is the cause of the unevenness. Thus it is doubtful that there is any existing statistical distribution currently widely used that would fit these curves.

The graph in Figure 3 shows the distribution of costs for all ages 20 through 60. Note that some of the "lumpiness" still remains. The right hand portion of the graph is of greatest concern to excess reinsurers and it is important to test the assumptions that go into the creation of this tail.

The medical costs are assumed to be distributed as follows

Fatal Medical
Distribution

Amount	Probability
$0	99.0%
$100,000	0.75%
$1,000,000	0.25%

When a joint distribution is created using the above distribution and the distribution of indemnity costs shown in Figure 3, the distribution shown in Figure 4 is obtained.

The distribution of loss and medical combined is presented in numerical form in Appendix C. The percent of costs excess $500,000 is found to be 43.08%. The interested reader with access to a spreadsheet should be able to duplicate these results.

III. Claim Characteristics, Details and Considerations

A. General Remarks

Previously the three types of claims that needed to be considered were discussed in very general terms. However, as noted earlier, changes in estimates that are small relative to ground up costs can be large with respect to Excess costs. Thus it is necessary to analyze the characteristics of these claims and contributions to the costs in a fair amount of detail. A detailed explanation of some of the cost characteristics and variation by state follows.

Figure #3

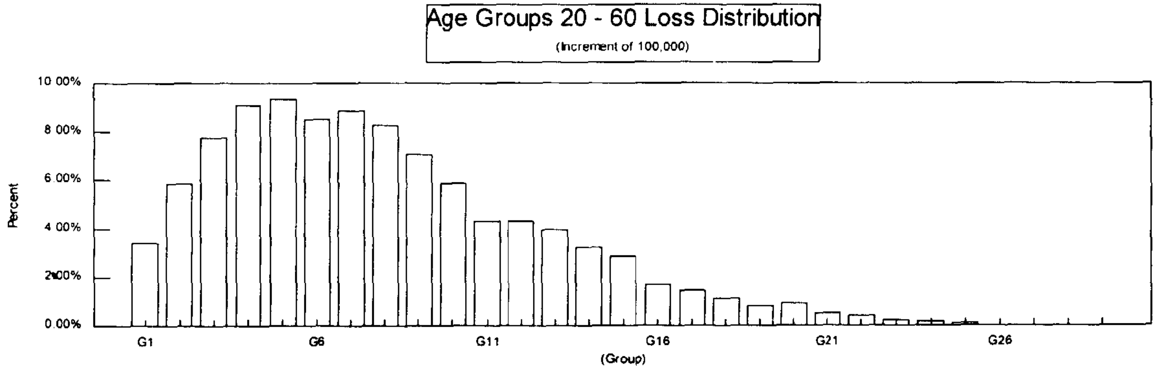
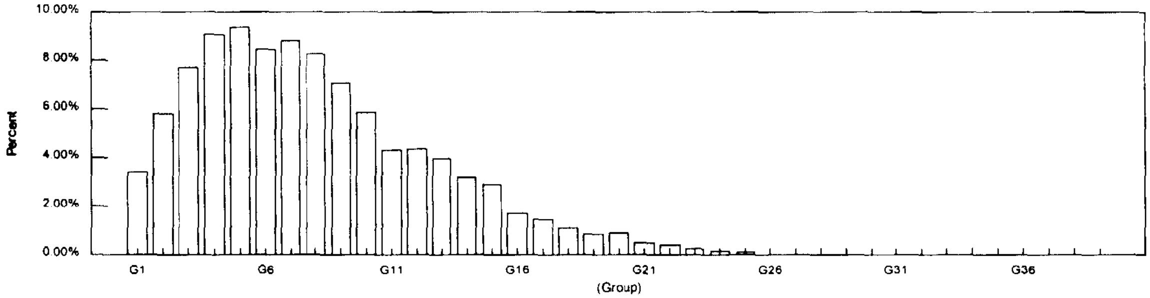

Age Groups 20 - 60 Loss Distribution
(Increment of 100,000)
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
Percent
G1
G6
G11
G16
G21
G26
(Group)

Figure #4

Loss & Medical Distribution

(Increment of 100,000)



B. Permanent Total

1. Medical

a. Comments/Range of Amounts

As mentioned previously it is difficult to obtain a single body of data that is sufficiently large and reasonably error free to be used in this process. Claims up to $20 million have been observed in the industry and it is thus reasonable to construct a distribution that accommodates claim costs of this amount even though the data on hand may not contain a claim of that size.

On the other end of the spectrum, ultimate incurred medical amounts that are less than $25,000 have been observed. This is difficult to explain. However, it has been suggested the accidents that are disabling such as blinding might be one explanation. Another is that some states have customarily awarded permanent total status for what seems to be minimal injuries. An example of this is an actual case where permanent total disability was awarded for tendinitis of the elbow. The medical costs of treating an injury of this type would be expected to be nominal.

Intuitively, the data may not be satisfying but given that the same thing is shown in several data sets it is reasonable to accept the indications.

b. Data/Quality, Amount, Culling

Given some of the observations above, it was found necessary to thoroughly review data sets almost on a claim by claim basis to eliminate claims which for one reason or another seem to have been erroneously included. For example claims whose incurred medical was below a certain cutoff point as of a given time e.g. two years after the date of accident were excluded. Also claims that demonstrated incurred medical but no indemnity were excluded. Other filtering protocols were also employed that resulted in data set felt to be free of at least obvious errors.

c. Development

Having cleaned up the data as much as possible the next step taken was to project individual costs to ultimate. At this point the only type of costs under discussion are the medical incurred amounts. Data was drawn from a recent Pennsylvania Loss Cost filing was used to develop the estimates in the following table.

Period	Case Res. Devl. Factor	Period	Case Res. Devl. Factor	Period	Case Res. Devl. Factor
12 mos.	2.82	84 mos.	3.26	156 mos.	2.61
24 mos.	2.66	96 mos.	3.42	168 mos.	2.39
36 mos.	2.79	108 mos.	3.37	180 mos.	2.20
48 mos.	2.86	120 mos.	3.02	192 mos.	1.90
60 mos.	3.13	132 mos.	2.84	204 mos.	1.81
72 mos.	3.21	144 mos.	2.70		

These factors are applied to the case reserves on individual open claims where the factors are selected according to the accident year of the claim. For example, suppose the year in which the data is being analyzed is 1999 and the accident year is 1992. Also assume that the undeveloped medical incurred is $272,312 where paid = $118,705. Then the ultimate medical incurred is (3.26)(153,607)+118,705 = 619,464.

This method sometimes will produce ultimate values that seem unreasonable and in that case judgement may have to be employed to temper the results.

d. Trending

The next step is to trend the cost on individual claims up to the current date. A good source of data for this purpose that is easily accessible is the Bureau of Labor Statistics. The web site address is www.bls.org. The medical increases for the last 10 years have been in the 3+% range.

After bringing the costs up to current level the costs are then projected to the middle of the period for which the rate will be applicable. Use of a future trend factor of approximately 3.5% at the writing of this paper seems reasonable.

e. Statistical Modeling

In previous applications of this method it has been found that the data even after the previously described adjustments is not smooth enough over various intervals to be used immediately. In particular it is often the case that there are ranges of several million dollars where there are no claims. Conversely – but occurring less often – there are instances when a fairly narrow interval might include two or more fairly large claims. "Fairly large" as used in this context means over 5 million.

Because of the large range of values and the characteristics of various types of medical claims associated with Permanent Total claims there is no reason to expect that any known statistical distribution will describe the distribution of medical claims. This is especially true with respect to fitting a single curve over the whole range of values. Some fitting over limited ranges may seem workable but the benefits seem questionable.

One well known curve that initially seemed appealing was the log-normal curve. However when a goodness of fit test was used (Komolgorov-Smirnov) on a medium size set of data the results were found to be inconclusive. Later when testing on a much larger body of data it become clear that the test results were indicating that it was unlikely that the data was generated as a sample from a log-normal distribution.

The solution adopted was to simply use the data as a foundation for an empirical curve. Before final construction of the curve, smoothing was conducted over consecutive intervals. A facsimile of the final curve is shown in Figures 5a-f.

2. Indemnity

a. Sampling vs. Modeling?

As indicated previously the objective of modeling of this component of the claim cost is to avoid estimation based on sampling of claims. In the case of the medical portion the nature of the actual causative mechanism is unknown. Thus it is necessary to resort to samples. However, that is not necessary with respect to indemnity and it was felt that a model could be constructed to estimate the costs with the resulting estimates possessing significantly less error than estimates produced by a sampling procedure.

b. Use of Statutorily Mandated Benefits

(1) Variation by State

Indemnity benefits vary by state with parameters associated with each of the following items displaying differences from one state to another.

(a) Function of Average Weekly Wage of Injured Individual

Most states define that indemnity benefits as a percent of wages. E.g. Alabama-66 2/3%; Georgia-66 2/3%; Idaho-67%; and so forth. However other states use different measures. E.g. Connecticut-75% of after tax income; Iowa-80% of spendable earnings; Michigan-80% of spendable earnings; Maine-80% of after tax AWW, etc. In addition, there are a few other states with somewhat more complex rules.

(b) Maximum and Minimums

As in the above states display differences in the Maximum indemnity awards. Most states define the maximum in terms of the state average weekly wage. For example, Alabama-100% SAWW; Colorado-91%

SAWW; Florida-100% SAWW; Iowa-200% SAWW Mississippi-66 2/3% SAWW. The maximum for New York is a dollar amount = $400.

Minimum show similar variations. Alabama-27.5% SAWW; Idaho 45% SAWW; Illinois-50% SAWW; Louisiana 20% SAWW; Michigan 25% SAWW etc. It should be noted here that there seems to be more variations in the Minimum than the Maximums. Many states have dollar amount minimums.

(c) Limits

In addition to the specifications in (1) and (2) above some states have limits specified in either time and/or amounts. Usually when there are limits these are expressed in both time and amounts. For example, South Carolina-500 weeks, $241,735; Mississippi-450 weeks, $131,787. For the most part however, the benefits are granted for life, although some states have offsets and other types of limitations that are discussed in the next section.

(d) Offsets

Some states have introduced Offsets and this trend has continued into the present time. For example: Arkansas-Reduce PP 50% of non-employee portion of public/private funded retirement/pension plan of 65 years or older; Colorado-Social Security, unemployment compensation, an employer-paid pension plan; Michigan-Disability, unemployment compensation, pension, old age Social Security retirement; New Jersey-Social Security; Pennsylvania-unemployment compensation, Social Security Old Age and certain severance and pension payments.

These offsets can be difficult to evaluate due to the vagueness of summaries of the law which are to be found in the most often used reference documents. However each state has contacts that will attempt to answer the questions. It must be kept in mind though that sometimes the answers are not correct with the implication that more than one source should be used if possible.

(e) Escalation

In most states the amount of weekly indemnity payable is determined close to the time of injury and remains at that level as along as payments are made. However some states mandate escalating benefits during part or all of the payment period. For example Florida requires escalation at 5% per year for 10 years with the escalation being an arithmetic increase, rather than a geometric increase. Connecticut and

Massachusetts mandate escalation tied to the CPI but limited to 5% in Massachusetts. Nevada's benefits are increased by an amount equal to the change in the SAWW.

(2) Estimation of Parameters by State

(a) Wage/Benefit Distribution

As an example of the way the distribution of benefits is calculated, the following is presented. It's assumed in the state of interest that the minimum benefit is 20% of the SAWW and that the maximum is 100% of the SAWW. Wage distribution data was obtained from the NCCI and the following table created

Wage Distribution

Wage Group	Wtd Avg	Wage Dist.	Wage Group	Wtd Avg.	Wage Dist.
1	0.300	4.1%	13	0.943	4.2%
2	0.354	1.9%	14	0.995	3.8%
3	0.411	2.8%	15	1.043	3.5%
4	0.460	4.3%	16	1.092	3.3%
5	0.518	4.6%	17	1.140	2.9%
6	0.566	4.8%	18	1.210	2.8%
7	0.625	5.3%	19	1.249	2.6%
8	0.672	5.1%	20	1.310	2.5%
9	0.732	5.0%	21	1.352	2.3%
10	0.783	4.9%	22	1.410	2.0%
11	0.836	4.8%	23	1.454	1.9%
12	0.891	4.6%	24	1.500	16.0%

If it is assumed that the SAWW is 600 and the benefit 66 2/3% times AWW, then the figures in the Wtd. Avg. column should be multiplied by 400 producing the following table.

Benefit Distribution

Wage Group	Wtd Avg	Wage Dist.	Wage Group	Wtd Avg	Wage Dist.
1	$120.00	4.1%	13	$377.20	4.2%
2	141.60	1.9%	14	398.00	3.8%
3	164.60	2.8%	15	417.20	3.5%
4	184.00	4.3%	16	436.80	3.3%
5	207.20	4.6%	17	456.00	2.9%
6	226.40	4.8%	18	484.00	2.8%
7	250.00	5.3%	19	499.60	2.6%
8	268.80	5.1%	20	524.00	2.5%
9	292.80	5.0%	21	540.80	2.3%
10	313.20	4.9%	22	564.00	2.0%
11	334.40	4.8%	23	581.60	1.9%
12	356.40	4.6%	24	600.00	16.0%

This is one of the building blocks of the excess costs.

(b) Age

It would obviously be very cumbersome to calculate the benefits by wage group across all working ages and then to compound the amounts with amounts from a medical distribution whose approximate range is 0-20 million. On the other hand it would be a mistake to oversimplify and, perhaps, chose as an average age of all workers, say 40 years.

The protocol outlined on this paper is to assume that some workers are age 20 at time of injury, some 25 and so forth in five-year intervals up to age 60. This makes the number of ages more manageable and it seems, through some research and analysis, still provides a good estimate of the costs.

(c) Life Tables

The life tables used in these calculations are the tables from the 1979-1981 experience period and is total population. Thus, it includes males, females and all races. This is obtainable from the Center for Disease Control and can be downloaded from their website.

Theses tables are used based on the assumption that the U.S. work force has the same proportions of men and women as does the general population. Another assumption implicitly made here is that men and women have equal exposure to serious injury.

It could and has been argued extensively that for Permanent Total injuries, or at least certain subsets, an impaired life table should be used. However medical care today has advanced to the point that even very seriously injured individuals can expect a normal life span. The NCCI undertook a study of impaired lives fairly recently (within the last 10 years) and published a life table based on the study. Review of that table did not offer convincing evidence that other than the total U. S. population should be used

(d) Offsets

It should be noted that the Florida benefit law is being used as an example here in the discussion of "offsets" and the analysis of these benefits should not be construed as applying to any other state. The law of each state must be analyzed on its own. Under Florida law the sum of benefits from both Social Security and the State (Workers Comp) is limited to 80% of ACE (Adjusted Current Earnings). It is assumed in this example that the individual is earning $475.00 per week. Also assume for the sake of specificity that the year of the accident is the year 2000. A simplifying assumption used at this point is that the ACE for this individual in the year 2000 is $475.

Next the Social Security benefits for the disabled workers must be estimated. The benefits are based on earnings through the previous year and hence the earnings are adjusted back to 1999. We then estimate the Social Security benefit based on that number and using the Social Security benefit structure. (This can be obtained form the Social Security benefit website). In this case, the Social Security benefits are found to be $210.34 per week.

The next step is to estimate the benefit under Workers Comp. Since the individual is earning $475 per week and the benefit is awarded at 2/3 AWW the benefit is $316.67 per week. The sum of the Social Security Benefit and the Workers Comp benefit is $527.01 which exceeds 80% of $475 by $147.01. This amount is the "Offset". Thus the Workers Comp benefits are reduced from $316.67 per week to $169.66. This leaves the sum at $380.00 = (.8)(ACE).

It should also be noted at this point that Florida provides for escalating benefits for a period. The interpretation of this part of the law made here is that the 5% increase applies to the amount $169.66.

The law in Florida operates in the above described ways since by agreement with the Social Security Administration, Social Security "pays first". It should be noted that the agreements have been worked out on a state-by-state basis.

(e) **State Average Weekly Wage**

Since rates are made to be effective for some period in the future historical information must be trended to that period. When a history of State Average Weekly Wage is available, this is used to trend to the rate effective period. An example of this is given in the following table.

Statewide Average Weekly Wage
-Maximum Weekly Benefit-
(Massachusetts)

10/1/97-9/30/98	$665.55
10/1/96-9/30/97	$631.03
10/1/95-9/30/96	$604.03
10/1/94-9/30/95	$585.66
10/1/93-9/30/94	$565.94
10/1/92-9/30/93	$543.30
10/1/91-9/30/92	$515.52

This is taken from the Commerce Clearing House publication "Workers Compensation".

If this type of information is not available, the NCCI Statistical Bulletin can be used in conjunction with wage increase information obtainable form the Bureau of Labor Statistics.

C. Fatal

1. Medical

As noted earlier most Fatal claims do not have any medical cost associated with them. However some Fatal claims do display medical costs in small, medium or even large amounts. The average cost of the medical on Fatal claims is very, very small in comparison to the indemnity costs. However the task here is to estimate the Excess costs and thus the medical costs although small in relation to first dollar costs can add significantly to Excess costs. This is especially true when the Fatal benefits are extremely limited as in Florida where Fatal benefits are limited to $100,000.

The distribution generated for use in this methodology looks something like the following

Fatal Medical
Distribution

Amount	Probability
$0	95.0%
$500	4.0%
$100,000	0.7%
$1,500,000	0.3%

2. Indemnity

As in the case of Permanent Total injuries it was decided during the development of this methodology that the best estimate of the distribution of indemnity costs incurred on Fatal claims would be produced by starting with an analysis of the Statutory mandated benefits. A discussion of the components of these awards following.

a. Age

It assumed here that the ages of workers was uniformly distributed and that the propensity to suffer a fatality was the same at each age. It must be noted here that this is a simplifying assumption. There is some data available that would indicate that the frequency of mortality is slightly higher for workers in their twenties than for workers at higher ages. It has been speculated that this is a result of young workers either not having been fully trained in safety procedures, simply lacking experience, being either more inclined to take risks or being less careful. It should be mentioned here that similar data indicates that workers between fifty and sixty are more inclined to suffer permanent total injuries than younger workers. In this case it has been speculated that older workers are simply less physically fit than younger workers with the following implications. The first is that the execution of a particular task is more likely to result in an injury to an older worker than a younger worker e.g. lifting an object weighing 70 pounds. The second is that, given a particular injury, it may be that a younger worker would have a propensity to heal more quickly and completely than an older worker. These considerations have not been incorporated into the model due to the lack of a highly reliable database.

However the actuary should make an effort to be aware of this and other types of information which are difficult to quantify but which would affect the underlying risk. This naturally should be communicated to any underwriter with whom the actuary might be working on this type of risk.

In order to perform the calculations it is necessary to assume a certain age or potential ages of the deceased worker. As is the case with Permanent Total claims discussed previously the assumption used here is that the worker's age at time of death was either 20, 25, ..., up to 60.

b. Wage

The starting point in determining the benefits to the primary dependent (usually the spouse) of a deceased worker is usually weekly wages (occasionally "spendable income" or "after tax income"). About 40% of the state have limitations in either time or amount. For the rest, the cost of the benefits are estimated by first constructing a distribution of wages in a given state. Use of the same method as outlined in Section II B 2b. (2)(a) [Permanent Total Indemnity] can be used here.

c. Percent Award

There are a variety of benefit awards depending upon whether or not there is a spouse, whether there are "school age" children and upon the existence and dependency of others, e.g. parents, siblings. The "school age" above is in quotes since the specific maximum age for a school age child varies by state except for a few states where there is actually no age limit. For example in New York the Percent of Wages is
a) Spouse Plus Children – 66 2/3,
b) Spouse Only –66 2/3,
c) One Child Only-66 2/3
while in Oklahoma the Percent of Wages is
a) Spouse Plus Children – 100%,
b) Spouse Only – 70%,
c) One Child Only – 50%

In addition there are variations such as the spouse's percentage increasing to a higher number after the children have left school. There are also lump sum payments to the spouse and/or children as well as funeral expenses and burial expenses.

Given the above it seems reasonable to select a conservative but uncomplicated approximate level of benefits. For example, in the case of Oklahoma cited above if we assume that at the time of death there is a surviving spouse and two children aged 9 and 12 then the payments are 100% for 11 years (23 maximum if in school), 85% for the next 3 years and 70% thereafter. In order to simplify the calculations, it seems reasonable to simply assume level payments at 80% for life.

d. Maximum, Minimum

Weekly fatality benefits are limited as is the case for Permanent Total. Usually the maximum and minimum can be shown to be a function of the SAWW. However benefits to children may cause some small exceptions. These limitations play a significant role when the benefits are payable for life but are not nearly as important when there are time or amount limitations. For example consider Florida where the limitation on Fatal benefits is $100,000.

The maximum weekly benefit in Florida is $522 per week. Thus the length of payments is about 3.7 years. If the weekly maximum was 50% higher the length of the payments would be about 2.5 years and if 50% lower, the length would be about 5.5 years. Thus the average point of payment would be either 1.85 years, 1.25 years or 2.75 years with the difference between any of these being no more than a year and a half. This is insignificant from the point of view of the time value of money and for excess rating purposes.

e. Offset, Limitation

(1) Offsets

Some states have Social Security offsets. Examples are Connecticut, District of Columbia, New York and Utah. The offsets for Fatal benefits are generally somewhat more complex that the offsets for Permanent Total Benefits. For example the New York law specifies "Where the death occurs on or after January 1, 1978 and the spouse is receiving benefits under the social security act for each $10 of the deceased average weekly wage in excess of $100, but in no case may the reduction exceed 50 percent of the spouse's share of the social security benefits.

- Average weekly wage over $100 up to and including $110, five percent;
- Average weekly wage over $110 up to and including $120, ten percent;

•
•
•

- Average weekly wage over $190 up to and including $190, forty five percent;
- Average weekly wage over $200 up to and including $110, fifty percent;

(2) Limitations

As noted previously about 40% of the states have some sort of aggregate limitation on the amount of Fatal benefit payments. For example California's limit on Fatal benefits is $125,000 ($160,000 if children), Florida's limit is $100,000, the Kansas limit is $200,000 and Texas limit $206,000. Other states have limits expressed in weeks. For example the Georgia limit is 400 weeks (or to age 65); the Idaho limit is 500 weeks; Illinois' limit is 20 years at TT rate whichever is greater), Virginia is 500 weeks and so forth.

It should be noted that any or all of these limitations can change in any year. Anyone employing the outlined method should consult the law or summaries of the law in specific states to determine the most current statutory limitations on benefits.

Escalation

A few states still mandate escalating benefits for fatal claim benefits. Most of these are in the Northeast e.g. Connecticut, Massachusetts and Rhode Island. It goes almost without saying that escalation is a major component costs. States with this type of benefit will exhibit the highest excess cost for workers compensation.

g. Mortality Table

As is the case with Permanent Total claims the life tables used in these calculations are the tables from the 1979-1981 U.S. experience and is derived from total population statistics. The implicit assumption made here is that men and women suffer fatalities equally in the workplace. This is probably not a precisely correct assumption and it has been speculated that perhaps the mortality rate is higher for men since men engage in inherently more hazardous work e.g. contracting, roofing, logging and fishing. However a considerable number of women drive or ride in vehicles as part of the job and many of the fatalities experienced in the course of work result from vehicle accidents. Whatever the true exposure, the unavailability of good data makes attempts to measure the mix of male and females with respect to fatal claims somewhat impractical. It should also be noted that use of an "all lives" mortality table when most of the workers compensation. fatalities are men adds a degree of conservatism. It might be noted here that in developing this methodology many similar decision points were encountered and the decision was made to make conservative selections due to the large degree of risk taken in underwriting an excess workers comp program.

Finally it should be noted that spouse's benefits generally cease upon remarriage and that a lump sum benefit is paid at this point. In an ideal world this might reduce the costs of the benefits somewhat. However because of changing options available to survivors, availability of remunerative work and other considerations and because of the lack of availability of reliable remarriage tables it was decided to ignore this feature of the workers comp benefit laws and thus add another bit of conservatism to the estimate.

D. Major Permanent Partial

1. Medical

a. Source of Data

Data on Major Permanent Partial claims can be obtained from the NCCI, actuarial consulting firms, large primary carriers with substantial books of workers compensation or perhaps other rating bureaus such as the PCRB or the NYCIRB.

Large established casualty reinsurers usually also have substantial databases on excess workers comp losses that they usually regard as proprietary. However reinsurer databases often suffer from two problems. One is that retentions have shifted dramatically over the years with the result that it is difficult to combine data from various years. In addition, information on serious claims that only presents part of the picture can be misleading. That is, trying to estimate the distribution of all Major Permanent Partial claims from a group of claims that have pierced a particular retention is significantly more difficult than working with the totality of this type of claim. And in order to use the methodology outlined in this paper, the entire distribution is needed.

b. Range of Amounts

It was mentioned earlier that the range of the medical costs associated with this type of injury can be surprisingly large. Some databases that we were able to access displayed claims whose maximum incurred medical was not much over $500,000. But other databases presented claims in the multiple millions of dollars. Serious injuries such as damage to the spinal column, severe burns requiring extensive reconstructive surgery and electrical burns causing nerve and muscle damage are only a few of the examples of medical catastrophes that are very costly but which may allow an individual to return to work.

The probabilities of this type of event are low as might be expected, but in constructing the distribution curve for medical costs on Major Permanent Partial claims consideration should be given to claims that could be in excess of $5 million. At the other end of the scale it's reasonable to assume that the medical cost of a Major Permanent Partial claim should have a minimum on the order of $15,000-$20,000.

The expected value of the average medical claim for Major Permanent Partial has been estimated to be between $80,000 and $100,000 in PCRB filings in recent years.

2. Indemnity

   a. Source of Data (not Statutory)

For Fatal and Permanent Total Claims it was felt that direct recourse to state Statutes would generate the best available estimate of indemnity costs associated with these types of claims. However this is not true with respect to the benefits provided for Major Permanent Partial. For one thing there are an inordinate number of categories e.g. loss of index finger, thumb, eye, great toe, other than great toe, foot, arm, hand, leg and on and on and on.

If a good current distribution of these injures by type were available (this would require a lot of injuries in each category to be credible) and the distribution could be expected to be applicable to the period for which the rates are to be effective (working environments are changing rapidly, so this is questionable) then this approach might be feasible. However we thought that the best way to estimate the distribution was to access a database of claims.

Data on the indemnity costs associated with Major Permanent Partial claims can be obtained from some of the sources previously cited. However it should be noted that the benefits for and definition of Major Permanent Partial claims vary significantly from one state to another. As a small example of this, Illinois law specifies "The specific case of loss of both hands, both arms, or both feet, or both legs, or both eyes, or any part thereof, or the permanent and complete loss of use thereof, constitutes total and permanent disability". In other states some of these described injuries would be classified as Major Permanent Partial.

Because of the variation from state to state the methodology adopted here has been to acquire a sample of Major PP's from a single state, thereby obtaining, it is hoped, homogeneous data and constructing the indemnity distribution curves. When there is not recourse to additional data for all the states the curve is adjusted by reviewing the details of the Statutory PP indemnity benefits.

It should be noted that indemnity benefits can be unexpectedly large particularly in comparison with the schedule benefits listed in summaries of the Workers Compensation laws. This is the result of the fact that in a number of states Temporary Total benefits may be received for as much as 500 weeks and is not to be deemed a reduction to subsequently awarded Permanent Partial benefits.

b. Range of Amounts

Indemnity benefits for Major Permanent Partial claims seem to be relatively well contained in comparison to the benefits that might be experienced on Fatal or Permanent Total claims. Thus from the point of view the excess insurer or reinsurer indemnity benefits are not much of a threat to higher retentions. However, because Major Permanent Partial claims are serious claims it might be expected that the indemnity costs will not be trivial. In addition when these costs are combined with moderately high medical costs even relatively high retentions will be penetrated.

Since Major Permanent Partial claims are serious it might be expected that the minimum indemnity costs will be in the range of 15 thousand to 20 thousand dollars. The indemnity benefits would contemplate temporary total plus scheduled benefits. On the other end of the range it is entirely possible that the maximum indemnity benefits that might be observed would be in the interval of 500,000 to 750,000 dollars. It should be noted that in some states compensation for temporary disability is allowed in addition to scheduled benefits, in others temporary benefits are allowed with some limitations and in some the temporary benefits are deducted from the scheduled amount.

Pennsylvania rate filings show estimated indemnity benefits averaging between 140,000 and 160,000 dollars.

IV. Weighting Excess Factors

It may seem surprising but the determination of the weights by type of loss may be the weakest link in this methodology. Often the weighting must be based on data that is the summary of data on a handful of claims.

This is particularly true in states with small populations. Thus some judgement, intuition and just plain common sense must be used in selecting the weights when estimating XS rates for a given state.

A. Source of Data

Data is available from the various statistical gathering and ratemaking organizations. The National Council on Compensation Insurance is the most prominent of these and issues a widely distributed and used Statistical Bulletin each year. A facsimile of part of Exhibit X from the 1998 bulletin appears below.

Distribution of Incurred Losses
By Type By State

State	Policy Period	Fatal %	Permanent Total %	Permanent Partial %	Temporary Total %	Medical Only %
AL	-	1.8	5.2*	67.3	21.1	4.6
CA	-	1.3	7.5	75.6	8.7	6.9
LA	-	2.5*	4.3*	54.2	30.8	8.2
MA	-	3.1	5.2	56.1	31.8	3.8
NY	-	-	-	-	-	-
TX	-	3.7	5.4	56.2	27.8	6.9

The asterisk (*) indicates that the figure is based on less than 25 cases. Given this, it might be expected that the indicated weight is not especially accurate since the sample size is small and that the range of values of individual claims is quite large.

In addition to the above cited weakness, the 1998 Edition also did not display weights for several states. Some were large states, notably Ohio and Pennsylvania.

Similar weights can be extracted from the rate filings of other rating bureaus such as PCRB, NYCIRB and WCIRBC.

B. Development

In addition to noting the problem of sparse data, it is also necessary to recognize the fact that development may not be to a truly ultimate value. The following table is taken from a recent Pennsylvania Loss Cost filing.

All Policy Years

1. Experience as Reported

	Indem.	Med.	Total	Prcnt.
Death	1,413.4	144.4	1,557.8	1.9%
Perm T.	3,636.7	1,891.6	5,528.3	6.7%
Maj. Perm Pa	24,898.2	7,577.1	32,475.3	39.6%
Min Perm Pa	6,825.5	4,946.8	11,772.2	14.3%
Temp Total	12,709.9	12,613.8	25,323.7	30.9%
Med. Only	-	5,437.8	5,437.8	6.6%
			82,095.1	100.0%

2. Developed Experience

	Indem.	Med.	Total	Prcnt.
Death	1,634.9	319.3	1,954.2	1.8%
Perm T.	6,512.2	4,581.0	11,093.2	10.2%
Maj. Perm Pa	36,037.2	19,046.9	55,084.1	50.4%
Min Perm Pa	6,017.6	4,697.9	10,715.5	9.8%
Temp Total	11,974.5	12,438.1	24,412.6	22.4%
Med. Only	-	5,854.6	5,854.6	5.4%
			109,114.2	100.0%

Less mature data exhibits a greater change in the distribution of the type of loss as the following table shows

Latest Policy Year

1. Experience as Reported

	Indem.	Med.	Total	Prcnt.
Death	194.4	20.1	214.5	2.2%
Perm T.	116.7	230.0	346.7	3.6%
Maj Perm Pa	1,139.6	403.8	1,543.4	16.0%
Min Perm Pa	1,165.0	798.9	1,963.9	20.3%
Temp Total	2,363.8	2,292.9	4,656.7	48.1%
Med Only	-	947.4	947.4	9.8%
			9,672.6	100.0%

2. Developed Experience

	Indem.	Med.	Total	Prcnt.
Death	267.3	56.5	323.8	1.7%
Perm T.	1,032.7	805.0	1,837.7	9.7%
Maj Perm Pa	6,197.3	3,281.6	9,478.9	50.0%
Min Perm Pa	957.0	755.3	1,712.3	9.0%
Temp Total	2,191.0	2,255.9	4,446.9	23.5%
Med Only		1,157.0	1,157.0	6.1%
			18,956.6	100.0%

Of particular interest in the two above tables is the development of the percentages for Permanent Total and Major Permanent Partial. The percentage of the third type of loss, Fatal, is close to 2.0% at first reporting and at its projected ultimate. However the Permanent Total percent develops substantially and the Major Permanent Partial only slightly less.

It should be noted that the above figures are taken from a primary rate filing with the development terminated after a reasonable amount of time. However, experience with Permanent Total claims would suggest that the cost of this type of claim continues to develop over a period measured in decades. Thus the distribution percentage for PT in particular is likely on the low side even at what is construed to be ultimate for the purposes of the rate filing. Thus the selection of the weights requires some judgement. For example the Permanent Total column of the above constructed facsimile shows weights between 5.2% and 7.5%. The states displaying 5.2% as the weight for PT are Alabama and Massachusetts. However Massachusetts is a much higher benefit state than Alabama with not only a higher average weekly wage but also with escalating benefits to age 65. On the other hand the fatal benefits in Texas are about the same as in Louisiana, so it is difficult to justify the difference in weights shown in he table. Thus, when selecting weights, consideration must not only be given to whatever data is available but also to the state mandated benefits.

V. Examples

Presented below is an additional example of the method under discussion. This is considered to be a true Excess Workers Compensation example with data sources and calculations being very close to what has be previously discussed.

A. Example #3

In the following it is assumed that the medical distribution Permanent Total Claims is as displayed in Figures 5a-5f. The indemnity is

Figure 5a

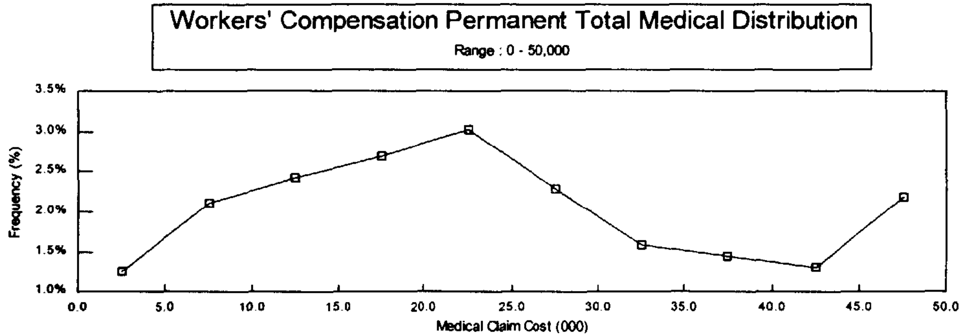


Figure 5b

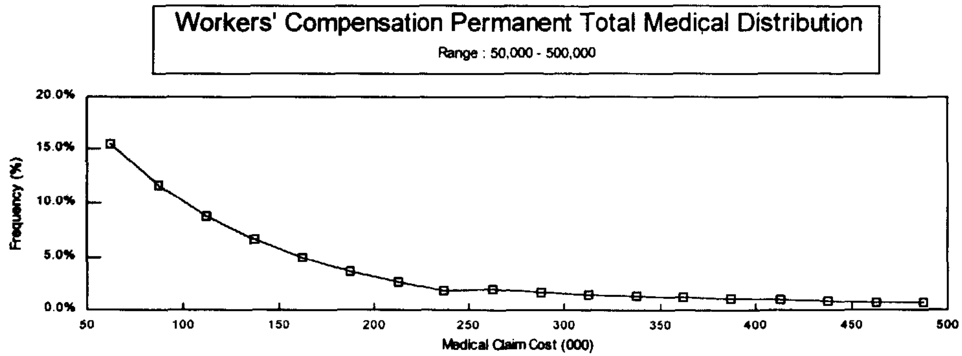


Figure 5c

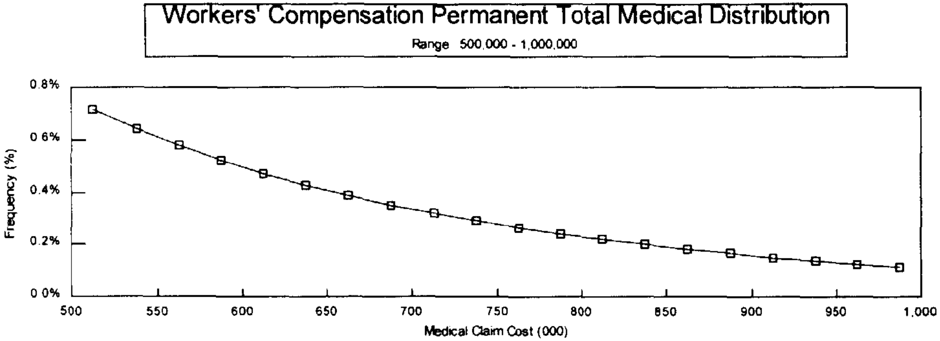


Figure 5d

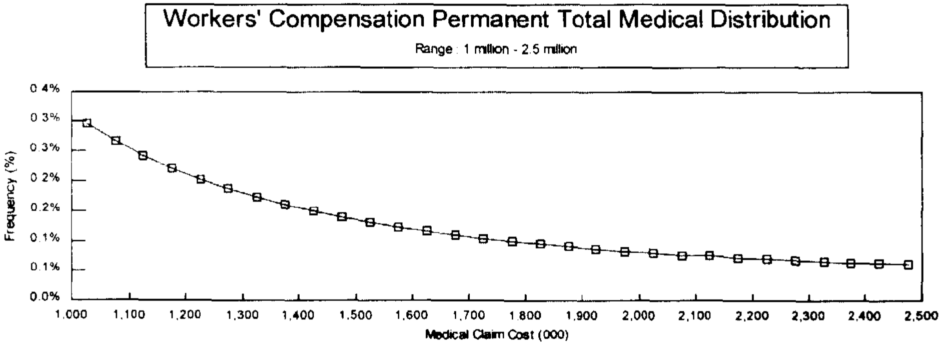


Figure 5e

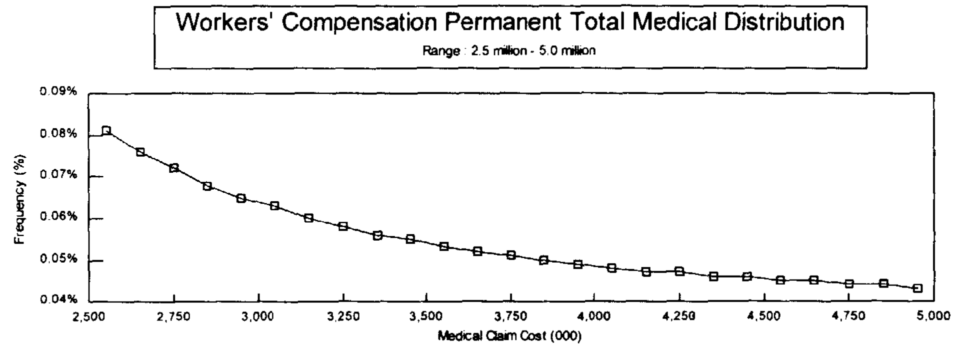
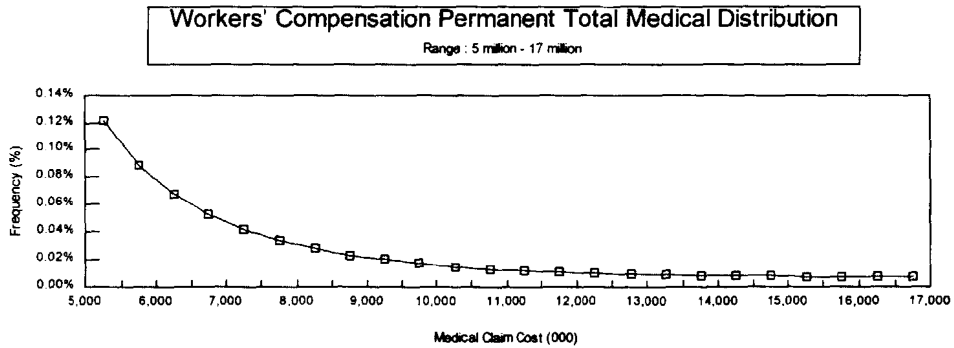


Figure 5f



generated by assuming a wage distribution similar to that produced by NCCI in the past and assuming a given level of SAWW and benefits. The SAWW is assumed to be $600 in this example.

The percent indemnity benefit is assumed to be 66 2/3% of wage at time of injury for Permanent Total claims. The maximum is 100% of the SAWW and the minimum is 20%. The self-insured retention (SIR) is $500,000. Given the above information the excess cost is found to be 55.8%.

For Fatal claims the medical distribution is as shown in the following table

Fatal
Medical Distribution

Amount	Probability
0	25.0%
8,000	67.5%
75,000	4.0%
300,000	3.0%
1,750,000	0.5%

Again the SAWW is assumed to be $600. The percent indemnity is assumed to be 50% for Fatal claims. The maximum and minimum percents are 100% and 20% respectively. The SIR's $500,000. The above assumptions result in an excess percentage of 34.1%.

Finally data with respect to Major Permanent Partials is displayed in Figures 6 and 7 following. Figures 6a, 6b and 6c display the medical costs. Figures 7a and 7b display the indemnity costs. The percent of costs excess $500,000 is 11.5%.

Next assume that the weights are as given in the following table.

Weights by
Type of Loss

Type of Loss	Weights
Fatal	2.0%
Perm. Total	11.5%
Maj. Perm. Pa.	55.0%

Figure 6a

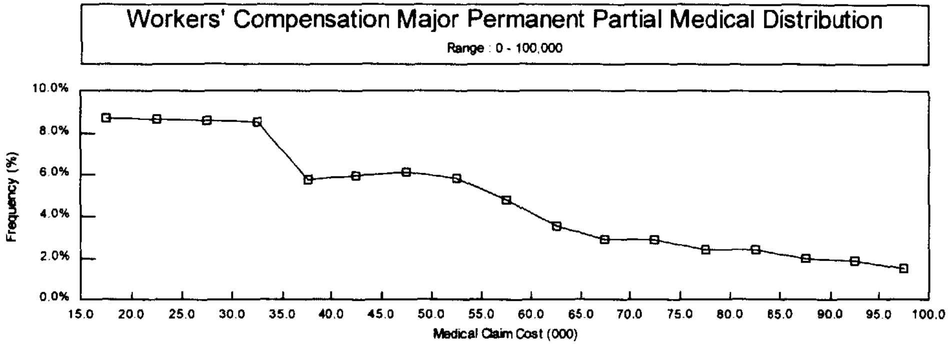


Figure 6b

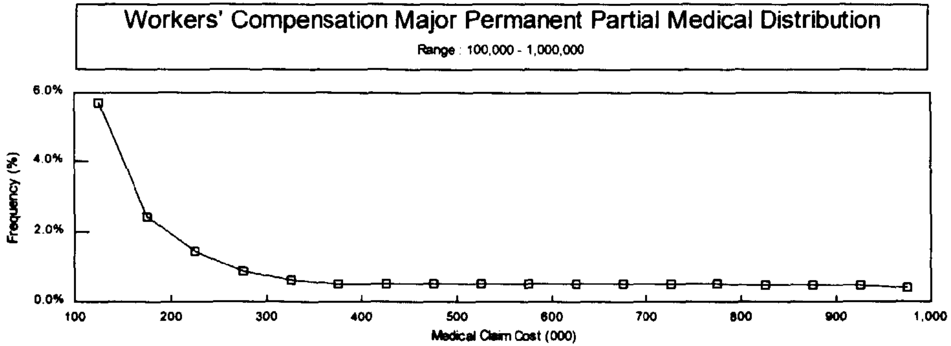


Figure 6c

Workers' Compensation Major Permanent Partial Medical Distribution

Range : 1,000,000 - 5,000,000

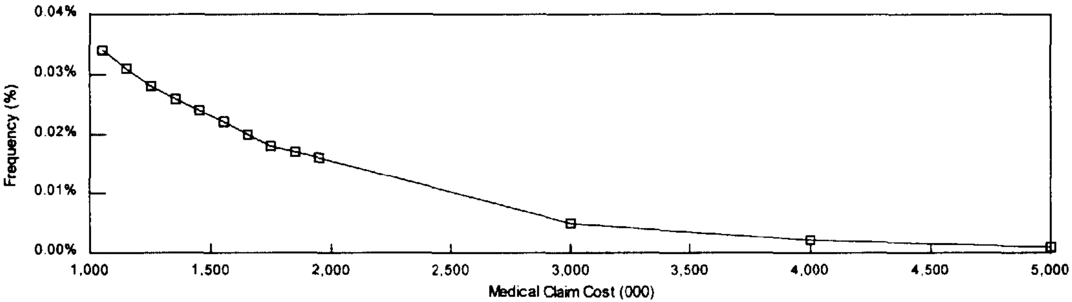


Figure 7a

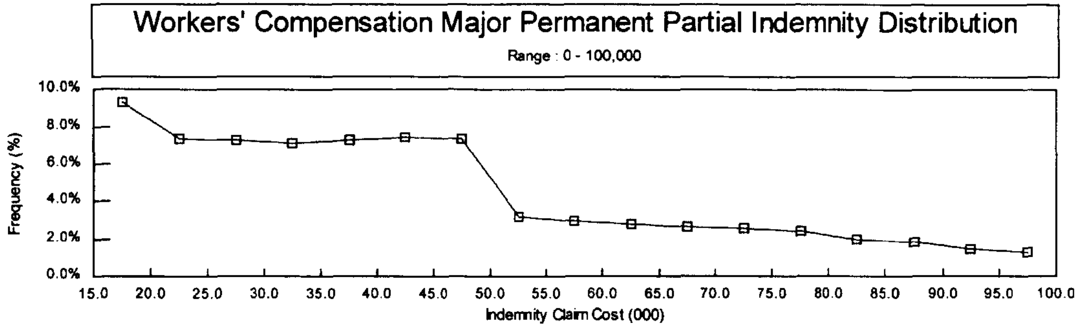
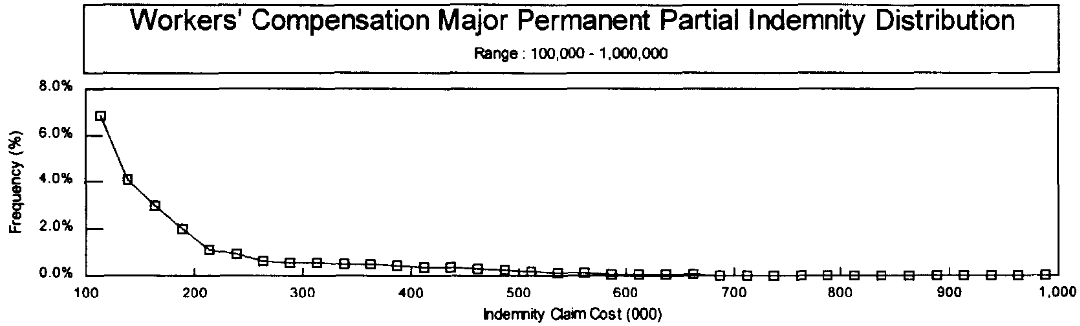


Figure 7b



These weights combine with the previously estimated excess factors to produce an excess factor of 13.45%, i.e.
(2.0%)(34.1%)+(11.5%)(55.8%)+(55.0%)(11.5%) = 13.4%.

B. Example #4

In Appendix D, the reader will find a complete set of excess factors for Pennsylvania. These were developed using the described methodology.

VI. Miscellaneous

1. Change of Benefits

Over the past decade workers compensation laws have been revised often with varying levels of impact emanating from given changes. Many of the changes have been focused on benefits. In an effort to bring the benefits accruing to an injured worker to a level equal to economic benefits accruing from other events, the benefits have generally been reduced and/or the administration of the law modified. For example, Maine at one time mandated escalating benefits for workers that had been killed or had been permanently and totally disabled. The benefits plus the rate regulation grew so onerous that eventually the insurance industry stopped underwriting workers compensation exposures in that state. The resulting problems that this caused businesses that operated in Maine were partially remedied by reducing the statutory benefits. Currently instead of escalating lifetime benefits fatal claims receive level benefits for 500 weeks. Permanent Total claim now receive level lifetime benefits but these are now offset to an extent by Social Security benefits and other benefits such as employer funded benefits.

Pennsylvania and Louisiana are two other states which have revised the statutorily mandated benefits in the last decade.

Changes such as these will naturally generate changes in excess costs, usually lowering them as a result of decreasing statutory benefits. Reliance on existing data to estimate the revised excess costs only makes sense if the particulars of previously incurred permanent total and fatal claims are known and present the possibility of estimating the costs under the new benefit system. Even when available this is tedious and expensive with the adjusted values being subject to some degree of error.

It is suggested that the methodology presented in this paper is superior in that an estimate of the excess factors that would be expected under the new law can be produced in a very short time. This not only saves expense which is usually somewhat important but it also saves time which is often more critical. In addition to the above, the view presented here is that the estimate generated using the described methodology is, at any rate, more accurate since it does not depend on a small sample of claims.

2. Uncertainty and Sensitivity Testing

   Even if all the steps described above are executed in a reasonably effective manner, there may still be a good deal of uncertainty in the final rates that are produced. One of the reasons for this is the data problem that has been discussed above a number of times. If there is one true statement that can be made about developing Excess Rates for Workers Compensation it's that there is never enough data available and that the ultimates on individual claims will never be known with sufficient accuracy to provide a great deal of comfort.

   Another of the reasons is that the law governing Statutory Benefits made not be entirely clear and/or interpretations of the law may be somewhat flawed. Finally, those administering the law may not be applying the law as intended.

   As a result of the uncertainty it is advisable to examine the rates that have been produced and to evaluate the contribution of each component. Once this is done the person charged with producing the final rates should test the sensitivity of the rates to changes in a given component or simultaneous changes in a number of components. This should provide a guide to which elements produce the greatest change in the rate for a given amount of error. Additional resources can then be brought to bear on the re-estimation for critical components.

3. Pricing of Layers

   The methodology presented here is designed to provide the cost of statutory benefits excess of a given retention. The cost is expressed as a percent of the pure premium. For example in Appendix D it can be seen that the cost excess $500,000 is 9.88%. However many specific excess treaties are written for layers such as $500,000 excess $500,000 (usually referred to as 500 xs 500). The excess pricing should be able to accommodate this. In Appendix D the price of the excess $1,000,000 layer is found to be 2.50%. Thus the cost of the layer 500 xs 500 is 7.38%.

   After constructing a table of excess factors a test of the results can be generated by examining the costs of consecutive layers such as 250 xs 250, 250 xs 500, 250 xs 750 etc. There of course should be no reversals and the costs should be decreasing uniformly. While this test does not guarantee that the results are accurate it is a simple task to perform and may identify missteps in calculations.

4. Cost Determination for "Compromise and Release" States

   Occasionally a question is raised with respect to whether the proposed method needs to be modified for use in states where serious claims can be settled essentially through paying lump sum to the injured party.

   The position adopted here is that it is necessary in constructing rates to provide for the costs of all claims regardless of how they are ultimately disposed. The first step is to estimate the nominal costs of all claims that will occur and estimate the cost of the various retention levels. The next step in converting these cost estimates into prices, is to estimate the impact of investment income. This leads to a price that is charged for the risk. In the event that a claim occurs, the funds plus future interest should be enough to pay for the claims or to pay the claims immediately on a present value basis. Thus those charging the calculated price in exchange for assuming the risk should be indifferent as to whether the claims are settled early or not.

5. Adjusting Statewide Indications to Reflect Individual Risk

   The methodology presented in this paper was designed to produce statewide rates. Thus the rates will be adequate but not excessive for a risk whose profile is exactly the same as the state as a whole. However the vast majority of risks presented to an underwriter will generate risk which is either greater or less.

   It has been suggested that adjusting the rates to reflect the Hazard Group profile might produce the appropriate rates. However it should be noted that somewhat over 90% of all risks fall into either Hazard Group II or Hazard Group III. Thus adjusting the statewide rates by Hazard Group may produce some improvement in matching the rates to the risk but it would seem that the progress would be minimal.

   It would seem that a better approach would be develop a profile of the risk by classification code with debits or credits assessed by code. The process of developing debits or credits by classification code is a major undertaking and is beyond the intended scope of this paper. However thinking along these lines will likely reproduce rates that more closely match the risk than recourse to Hazard Group.

   Statistics to begin the above suggested process are available from the various ratemaking bureaus.

6. Allocated Loss Adjustment Expense Considerations

   The methodology and the examples presented in this paper did not consider the impact of allocated loss adjustment expense. However it is felt that this methodology can be extended to include allocated loss adjustment expense costs. It would seem that this would add an additional layer of complexity. Evidence available to the author of this paper suggests that ALAE is not a direct add-on. That is, it would be inappropriate to load each claim value by say 10%. For example, a claim whose size is $15,000,000 would not carry an associated ALAE cost of $1,500,000.

   On the other hand, whereas the medical and indemnity costs seem to be independent, it would appear that the ALAE amount is, in some way, related to the size of the claim cost excluding ALAE. However incurred ALAE as a percent of incurred losses seems to be negatively correlated to the size of loss.

7. Payout Rates

   There may be an initial temptation to model the payout of the incurred claim costs evenly over a lifetime. This is generally incorrect. For limited benefit fatal claims, Major Permanent Partial and Employer's Liability the average date of payment is actually within three to seven years of the accident date.

   Permanent Total claims present something of a paradox especially in comparison to, say, General Liability. As a rule of thumb, the larger G.L. claims are paid later and hence it might be expected that additional investment income might be generated to offset the cost of the ultimate settlement (This is not an inviolate rule.) However, larger Permanent Total Claims are, all things being equal, caused by larger medical costs. Claims like these generally, (not always), demonstrate extremely large medical payments in the earlier years to counteract the effect of the serious injuries. Thus, generally, the larger the Permanent Total claim, the faster the payments. On PT claims where the incurred medical in very high (excess of 2-5 million) a retention of 500,000 can be pierced in a year or two.

8. Closing Remarks

   The method outlined in this paper was developed in response to a specific problem. The problem - generation of reasonably accurate estimates of excess workers compensation costs- is sufficiently important and of wide enough interest to justify the cost in some circumstances.

The effort, cost, and acceptance of the methodology do not guarantee, of course, that the rates are as accurate as they might be. This is due in part to the difficulties previously discussed. It is also due to assumptions that have been untested but where at least a degree of testing may be possible. Thus work must continue to refine the methodology.

Pennsylvania – Compensation Benefits
Summary of Salient Items

Death Benefits

Dependents: In the case of death, compensation will be computed on the following basis, and distributed to the following persons, provided that in no case will the wages of the deceased be taken to be less than 50 percent of the SAWW.

Children, no spouse: If there is no surviving spouse entitle to compensation, compensation will be paid to the guardian of the child or children, if there is no guardian, then to such other persons as may be designated by the board as follows

- If there is one child, 32 percent of wages of deceased, but not in excess of the SAWW
- If there are two children, 42 percent of wages of deceased, but not in excess of the SAWW
- If there are three children, 52 percent of wages of deceased, but not in excess of the SAWW
- If there are four children, 62 percent of wages of deceased, but not in excess of the SAWW
- If there are five children, 64 percent of wages of deceased, but not in excess of the SAWW
- If there are six children, 66 2/3 percent of wages of deceased, but not in excess of the SAWW

Spouse and children: To the widow or widower, if there is one child, 60 percent of wages, but not in excess of the SAWW. To the widow or widower, if there are two children, 66 2/3 percent of wages but not in excess of the SAWW. To the widow or widower, if there are three or more children 66 2/3 per cent of wages, but not in excess of the SAWW.

Parents: If there are neither widow, widower, nor children entitled to compensation, then to the father or mother, if dependent to any extent upon the employee at the time of the injury, 32 percent of wages but not in excess of the SAWW. (Additional wording omitted)

Siblings: If there are neither widow, widower, children, nor dependent parent, entitled to compensation, then to the brothers and sisters, if actually dependent on the employee for support at the time of death, 22 percent of wages for one brother or sister, and an additional five percent for each additional brother

R

or sister, with a maximum of 32 percent of the wages of the employee, not to exceed the SAWW.

Generally, compensation is payable to or on account of any child, brother, or sister, only if and while the child, brother, or sister, is under the age of 18. If the child, brother, or sister is dependent because of disability, then compensation will be paid during the disability of a child, brother, or sister over 18 years of age. Furthermore, if the child is enrolled as a full-time student in any accredited educational institution, then compensation will continue until the student turns 23. (Additional wording omitted)

Spouse Only: To a surviving spouse if there are no children, 59 percent of wages not to exceed the SAWW

Miscellaneous Benefits:

Funeral Expenses Whether or not there are dependents, the reasonable expense of burial, not exceeding $3,000 will be paid by the employer or insurer directly to the undertaker (without deduction of any amounts already paid for compensation or for medical expenses).

Permanent Disability Compensation

Permanent Total Disability: For total disability, 66 2/3 percent of the wages of the injured employee beginning after the seventh day of total disability, and payable for the duration of total disability. However, compensation cannot be more than the maximum compensation payable. If the benefit is less than 50 percent of the SAWW, the benefit payable will be the lower of 50 percent of the SAWW or 90 percent of the employee's average weekly wage. (Additional wording omitted)

Permanent Partial Disability: For partial disability, 66 2/3 percent of the difference between the wages of the injured employee before the injury and the earning power of the employee thereafter; but compensation cannot be more than the maximum compensation payable. (Additional wording omitted)

Schedule of Permanent Injuries. For all disability resulting from permanent injuries of the following classes, 66 2/3 percent of wages is exclusively paid for the following number of weeks

• loss of hand	335 weeks
• loss of forearm	370 weeks

R

• loss of an arm	410 weeks
• loss of a foot	250 weeks
• loss of a lower leg	350 weeks
• loss of a leg	410 weeks
• loss of an eye	275 weeks
• loss of a thumb	100 weeks
• loss of a first or index finger	50 weeks
• loss of a second finger	40 weeks
• loss of a third finger	30 weeks
• loss of a fourth or little finger	28 weeks
• loss of a great toe	40 weeks
• loss of any other toe	16 weeks

(Additional wording omitted – including lengthy section on Hearing Loss)

Healing period compensation: In addition to the payments provided for permanent injuries of the classes specified, any period of disability necessary and required as a healing period is compensated in accordance with the provisions of this subsection. The healing period ends when the claimant returns to employment without impairment in earnings, or on the last day of the period specified in the following table, whichever is the earlier.

• For the loss of hand	20 weeks
• For the loss of forearm	20 weeks
• For the loss of an arm	20 weeks
• For the loss of a foot	25 weeks
• For the loss of a lower leg	25 weeks
• For the loss of a leg	25 weeks
• For the loss of an eye	10 weeks
• For the loss of a thumb or part thereof	10 weeks
• For the loss of a any finger or part thereof	6 weeks
• For the loss of a great toe or part thereof	12 weeks
• For the loss of a any other toe or part thereof	six weeks

(Additional wording omitted)

R

79-81 U.S.Standard Life Table

Age	Number of lives	Age	Number of lives
0	100,000	56	87,551
1	98,740	57	86,695
2	98,648	58	85,776
3	98,584	59	84,789
4	98,535	60	83,726
5	98,495	61	82,581
6	98,459	62	81,348
7	98,426	63	80,024
8	98,396	64	78,609
9	98,370	65	77,107
10	98,347	66	75,520
11	98,328	67	73,846
12	98,309	68	72,082
13	98,285	69	70,218
14	98,248	70	68,248
15	98,196	71	66,165
16	98,129	72	63,972
17	98,047	73	61,673
18	97,953	74	59,279
19	97,851	75	56,799
20	97,741	76	54,239
21	97,623	77	51,599
22	97,499	78	48,878
23	97,370	79	46,071
24	97,240	80	43,180
25	97,110	81	40,208
26	96,982	82	37,172
27	96,856	83	34,095
28	96,730	84	31,012
29	96,604	85	27,960
30	96,477	86	24,961
31	96,350	87	22,038
32	96,220	88	19,235
33	96,088	89	16,598
34	95,951	90	14,154
35	95,808	91	11,908
36	95,655	92	9,863
37	95,492	93	8,032
38	95,317	94	6,424
39	95,129	95	5,043
40	94,926	96	3,884
41	94,706	97	2,939
42	94,465	98	2,185
43	94,201	99	1,598
44	93,913	100	1,150
45	93,599	101	815
46	93,256	102	570
47	92,882	103	393
48	92,472	104	267
49	92,021	105	179
50	91,526	106	119
51	90,986	107	78
52	90,402	108	51
53	89,771	109	33
54	89,087	110	21
55	88,348	111	0

Appendix C

Indemnity & Medical   Age : 20-60

Group	Probability	Amount
G1	3.392298%	58,942.96
G2	5.809995%	155,860.20
G3	7.702869%	255,115.73
G4	9.049774%	354,567.58
G5	9.333479%	455,459.36
G6	8.461007%	553,659.39
G7	8.817601%	649,715.82
G8	8.254022%	749,451.97
G9	7.031388%	850,871.57
G10	5.844955%	953,687.72
G11	4.316986%	1,052,420.61
G12	4.333955%	1,146,177.19
G13	3.974101%	1,247,102.48
G14	3.206946%	1,346,987.28
G15	2.891444%	1,449,632.62
G16	1.716948%	1,551,120.97
G17	1.479908%	1,645,467.42
G18	1.136829%	1,748,418.39
G19	0.849223%	1,843,873.16
G20	0.907306%	1,948,071.59
G21	0.525709%	2,055,019.66
G22	0.401405%	2,148,343.07
G23	0.251621%	2,243,637.95
G24	0.157137%	2,336,717.51
G25	0.097350%	2,440,360.63
G26	0.029220%	2,550,790.86
G27	0.012159%	2,644,844.76
G28	0.005467%	2,743,043.15
G29	0.003101%	2,838,520.33
G30	0.002246%	2,947,941.25
G31	0.001283%	3,055,167.92
G32	0.000977%	3,148,337.99
G33	0.000603%	3,243,437.87
G34	0.000372%	3,336,082.85
G35	0.000225%	3,439,668.49
G36	0.000061%	3,551,044.20
G37	0.000021%	3,644,484.67
G38	0.000007%	3,737,255.81
G39	0.000003%	3,827,854.55
	100.000000%	753,447.87

Group	Probability	Amount
G1	3.426564%	58,942.96
G2	5.842723%	155,846.50
G3	7.736413%	255,111.55
G4	9.082577%	354,564.07
G5	9.358949%	455,465.94
G6	8.475570%	553,644.28
G7	8.842459%	649,687.29
G8	8.270407%	749,450.07
G9	7.039757%	850,884.23
G10	5.850663%	953,713.27
G11	4.307616%	1,052,394.21
G12	4.330345%	1,146,097.40
G13	3.961901%	1,247,071.31
G14	3.186389%	1,346,931.95
G15	2.872877%	1,449,607.32
G16	1.691124%	1,551,108.52
G17	1.459716%	1,645,353.36
G18	1.116369%	1,748,429.45
G19	0.831567%	1,843,676.94
G20	0.895396%	1,948,009.42
G21	0.513358%	2,055,167.92
G22	0.390635%	2,148,337.99
G23	0.241198%	2,243,437.87
G24	0.148850%	2,336,082.85
G25	0.089951%	2,439,668.49
G26	0.024563%	2,551,044.20
G27	0.008410%	2,644,484.67
G28	0.002640%	2,737,255.81
G29	0.001013%	2,827,854.55
Total :	100.000000%	750,197.87

Appendix D

Sample rates constructed using described methodology

State: Pennsylvania
Effective Year: 1999

Excess of	Excess Factor
100,000	37.67%
150,000	30.51%
200,000	25.32%
250,000	21.44%
300,000	18.31%
350,000	15.68%
400,000	13.43%
450,000	11.51%
500,000	9.88%
750,000	4.97%
1,000,000	2.50%
1,250,000	1.55%
1,500,000	1.07%
2,000,000	0.66%

Surplus Allocation for the Internal Rate of Return Model: Resolving the Unresolved Issue

Daniel F. Gogol, Ph.D., ACAS

SURPLUS ALLOCATION FOR THE INTERNAL RATE OF RETURN MODEL: RESOLVING THE UNRESOLVED ISSUE

Abstract

In this paper, it is shown that with a certain definition of risk-based discounted loss reserves and a certain method of surplus allocation, there is an amount of premium for a contract which has the following properties:

(1.) It is the amount of premium required for the contract to neither help nor hurt the insurer's risk-return relation.

(2.) It produces an internal rate of return equal to the insurer's target return.

If the insurer gets more than this amount of premium, then the insurer can get more return with the same risk by increasing the percentage of the premium for the overall book which is in the segment. Conversely, if the insurer gets less than this amount of premium, the insurer can increase its return by decreasing the percentage of the overall premium which is in the segment. The amount of premium is equal to the risk-based premium in "Pricing to Optimize an Insurer's Risk-Return Relation," (PCAS 1996).

The above property 1 of risk-based premium is proven by Theorem 2 of the 1996 PCAS paper and not by the present paper. The present paper proves property 2.

1. INTRODUCTION

The problem of relating pricing to the risk-return relation has been discussed in many recent actuarial papers. Surplus allocation is not described in these papers as something that can be done in a theoretically justifiable way. Actually, though, surplus allocation has been used in a way that has been proven by a theorem (Theorem 2 of [1]) to derive the amount of premium for any contract which will neither improve nor worsen the insurer's risk-return relation. Certain estimates have to be used of course, e.g. covariances and expected losses. The precise mathematical relationship between this premium and the risk-return relation is specified by the statement of Theorem 2. This theorem, and the corresponding definition of risk-based premium, are very rarely mentioned in recent papers relating pricing to the risk-return relation.

Since the internal rate of return (IRR) model has been a part of the CAS exam syllabus for years, and since it is a widely used method in insurance and other industries, it may be possible to explain the method of [1] to a larger group of readers by relating it to the IRR method.

The IRR model can be used to measure the rate of return for an insurance contract or a segment of business, but only if the method used for allocating surplus can be related to the insurer's risk-return relation. The model is generally presented without a theoretically justifiable method of allocating surplus. But if an arbitrary method of allocation (such as allocating in proportion to expected losses) is used, the results are almost meaningless. The purpose of this paper is to complete the IRR method.

Incidentally, there are several actuarial papers which argue that surplus allocation doesn't make sense because risk is not additive, or because in the real world all of surplus is available to support all risks. Actually, just as a function f(x) associates each number x with another number, surplus allocation is a mathematical function which associates each member of a set of risks with a portion of surplus. This function can be used as a part of a chain of reasoning in order to prove a theorem, as was done in Theorem 2 of [1].

Although surplus allocation was used in deriving the properties of risk-based premium in [1], the derivation doesn't actually require any mention of surplus allocation. When risk-based premium is related to the IRR method in section 4 of this paper, surplus allocation is used because that is the traditional way of explaining the IRR method.

The approach presented here:

(1.) addresses the risk-return relation in a fundamental way

(2.) addresses the problems of the time value of money and the discounting of losses

(3.) addresses the problem of loss reserve risk

(4.) assigns a risk-based premium to the sum of two contracts or segments which equals the sum of the individual risk-based premiums

The premium derived in this paper by the IRR method is the same as that determined by the method in [1] The method in [1] is simpler to apply, but the IRR model has the advantage of being widely used and understood. It is on the CAS syllabus and has also been used by non-actuaries for many years. In order to relate the method in [1] to the IRR method, explanations will be given of both methods. However, since both methods are explained at length in the literature (see [2],[3],[4]), the explanations will be brief and informal. This could actually be an advantage, since it could make the presentation more lively and readable. The part of the paper which is new is the derivation of the equivalence of the two methods, given certain assumptions and conditions.

2. THE IRR MODEL

The IRR model is a method of estimating the rate of return from the point of view of the suppliers of surplus. Suppose for example that an investor supplied \$100 million to establish a new insurer, and that \$200 million in premium was written the next day. Also suppose that twenty years later the insurer was sold for \$800 million. Ignoring taxes, the return r to the investor satisfies the equation \$100 million $(1+r)^{20}$ = \$800 million. Therefore, the return r equals 10.96%.

Suppose that, beginning at the time of the above initial investment, each dollar of surplus is thought of as being assigned to either an insurance policy currently in effect, a loss reserve liability, or some other risk. Suppose that a cash flow consisting of premium, losses, expenses, outflows of surplus, and inflows of surplus, is assigned to each policy in such a way that the following is satisfied: the total of all the cash flows minus the outstanding liabilities immediately prior to the time at which the above insurer is sold for \$800 million produces an \$800 million surplus. It is then possible to express the input of \$100 million, and the payback of \$800 million twenty years later, as the total result of the individual cash flows assigned to each policy. Based on the individual cash flows, the overall return of 10.96% could then be expressed as a weighted average of individual returns for each policy. The individual return for a policy is called its internal rate of return. The following example is taken from [2].

<u>An Equity Flow Illustration</u>

A simplified illustration of an insurance internal rate of return model should clarify the relationships between premium, loss, investment, and equity flows. There are no taxes or expenses in this heuristic example. Actual Internal Rate of Return models, of course, must realistically mirror all cash flows.

Suppose an insurer

- collects $1,000 of premium on January 1, 1989,
- pays two claims of $500 each on January 1, 1990, and January 1, 1991,
- wants a 2:1 ratio of undiscounted reserves to surplus, and
- earns 10% on its financial investments.

The internal rate of return analysis models the cash flows to and from investors. The cash transactions among the insurer, its policyholders, claimants, financial markets, and taxing authorities are relevant only in so far as they affect the cash flows to and from investors.

Reviewing each of these transactions should clarify the equity flows. On January 1, 1989, the insurer collects $1,000 in premium and sets up a $1,000 reserve, first as an unearned premium reserve and then as a loss reserve. Since the insurer desires a 2:1 reserves to surplus ratio, equity holders must supply $500 of surplus. The combined $1,500 is invested in the capital markets (e.g., stocks or bonds).

At 10% per annum interest, the $1,500 in financial assets earns $150 during 1989, for a total of $1,650 on December 31, 1989. On January 1, 1990, the insurer pays $500 in losses, reducing the loss reserve from $1,000 to $500, so the required surplus is now $250.

The $500 paid loss reduces the assets from $1,650 to $1,150. Assets of $500 must be kept for the second anticipated loss payment, and $250 must be held as surplus. This leaves $400 that can be returned to the equity holders. Similar analysis leads to the $325 cash flow to the equity holders on January 1, 1991.

Thus, the investors supplied \$500 on 1/1/89, and received \$400 on 1/1/90 and \$325 on 1/1/91. Solving the following equation for v

$$\$500 = (\$400)(v) + (\$325)(v^2)$$

yields v = 0.769, or r = 30%. (V is the discount factor and r is the annual interest rate, so v = 1/(1+r).)

The internal rate of return to investors is 30%. If the cost of equity capital is less than 30%, the insurer has a financial incentive to write the policy.

This concludes Feldblum's example. My only attempt to improve this simplified illustration is the following. The \$1,000 reserve set up on January 1, 1989 is an unearned premium reserve and by the end of 1989 there is a \$1,000 loss reserve. In between, the sum of the unearned premium reserve and the loss reserve is always \$1,000.

Feldblum doesn't claim that the method of surplus allocation in the illustration can be directly related to the risk-return relation. Allocating in proportion to expected losses doesn't distinguish between the riskiness of unearned premium, loss reserves, property risks, casualty risks, catastrophe covers, excess layers, and ground-up layers, for example. Different methods of surplus allocation could be judgmentally applied to different types of contracts, but from a theoretical risk-return perspective a certain use of covariance is required. This will be explained in the next section.

3. RISK-BASED PREMIUM

What follows is an informal explanation of the derivation in [1] of the properties of risk-based premium. In the discussions below of an insurer's risk-return relation over a one year time period, "return" refers to the increase in surplus, using the definition of surplus below. (The term "risk-based discounted" is used in the definition and will be explained later.)

surplus = market value of assets - risk-based discounted loss and loss adjustment reserves – market value of other liabilities (3.1)

At any given time, the return in the coming year is a random variable. The variance of this random variable is what we refer to by the term "risk". The expression "optimizing the risk-return relation" is used in the same way that Markowitz [5] used it, i.e., maximizing return with a given risk or minimizing risk with a given return. (Markowitz was awarded the Nobel Prize several years ago for his work on optimizing the risk-return relation of asset portfolios.)

For an insurance contract, or for a segment of business, the risk-based premium can be expressed as follows: (The term "loss" will be used for "loss and loss adjustment expense.")

risk-based premium = expense provision + risk-based discounted losses + risk-based profit margin. (3.2)

The above expense provision is equal to expected expenses discounted at a risk-free rate. The starting time T for discounting recognizes the delay in premium collection. Expenses are considered to be predictable enough so that the risk-free rate is appropriate. The risk-based discounted loss provision is equal to the sum of the discounted values, using a risk-free rate and the above time T, of

(a.) the expected loss payout during the year

(b.) the expected discounted loss reserve at year-end, discounted as of year-end at a "risk-based" (not "risk-free") discount rate

A risk-free rate is used to discount (a) and (b) above because the risk arising from the fact that (a) and (b) may differ from the actual results is theoretically correctly compensated by the risk-based profit margin (see (3.2) above).

The phrase "contract or segment of business" will be replaced below by "contract", since the covariance method used below has the following property: the risk-based premium for a segment equals the sum of the risk-based premiums of the contracts in the segment. At the inception of an insurance policy, the payout of losses during the year that the contract is effective, and the estimated risk-based discounted loss reserve for the contract at the end of the year, are unknown. The effect of the contract on surplus at the end of the year, i.e., the difference in end of year surplus with and without the contract, can be thought of as a random variable X at inception. The insurer's return, i.e., the increase in surplus during the year, is also a random variable. Call it Y.

Assuming that the contract premium equals the risk-based premium, the expected effect of the contract on surplus at the end of the year is equal to the accumulated value, at the risk-free interest rate, of the risk-based profit margin. This is true because the expense provision portion of the formula (3.2) above pays the expenses, and the risk-based discounted losses portion pays the losses during the year and also accumulates at risk-free interest, to the expected value of the risk-based discounted loss reserves at the end of the year. Therefore, by formulas (3.1) and (3.2), above, the effect of the contract equals the accumulated value of the risk-based profit margin.

The random variables X and Y were defined above. If

$$Cov(X,Y)/Var(Y)=E(X)/E(Y) \tag{3.3}$$

then, according to Theorem 2 of [1], the contract neither improves nor worsens the risk-return relation, in a certain sense. This was defined above in the abstract. Note that E(X), above, equals the accumulated value of the risk-based profit margin.

One of the components of risk-based premium is the expected value of the risk-based discounted loss reserves at the end of the year. This expected value is greater than the expected value of the loss reserves discounted at the risk-free rate corresponding to the duration of the loss reserves. This is how the risk-based premium provides a reward for the risk of loss reserve variability. The risk-based discount rate is therefore less than the risk-free rate.

At the end of each year following the effective period of a contract, if the matching assets for the risk-based discounted loss reserves are invested at the risk-free rate, their expected value at the end of the following year will be greater than the expected discounted liability. This is because the risk-based discount rate is less than the risk-free rate. Assume, for example, that the loss payout is exactly equal to the expected loss payout. At the moments that loss payments are made, both the discounted loss reserve and the matching assets are reduced by the same amount. At other times, the matching assets are growing at the risk-free rate and the discounted liability is growing at the lower risk-based discount rate.

At the beginning of the second year after the inception of the policy, the end of the year matching assets minus the discounted loss reserve can be thought of as a random variable Z. If Cov(Y,Z)/Var(Y) is equal to E(Z)/E(Y), then, according to Theorem 2 of [1], the risk-based discounted loss reserve and matching assets neither improves nor worsens the risk-return relation for the year. It is possible to compute a discount rate before the inception of the contract such that Cov(Y,Z)/Var(Y) is equal to E(Z)/E(Y). Note that if the matching assets are not risk-free, that affects both Cov(Y,Z) and E(Z) and may have a slight effect on the risk-based discount rate. If risk-based discount rates are computed for each of the years until the loss reserve is expected to be fully paid, the risk-based premium for the contract is determined.

For practical purposes, the above derivation of risk-based discounting of losses for a contract can be simplified if certain estimates are used. For example a single risk-based discount rate can be used for all future years. This approach was used in [1]. Since risk-based premium is determined by the estimated expense payout, loss payout, risk-based profit margin, and risk-based discount rate, the explanation of risk-based premium has now been concluded. The following two examples are taken from [1].

Catastrophe Cover Risk Load

In this example, in order to estimate the value of a catastrophe cover to a ceding company, we will suppose that the ceding company re-assumes the cover, and we will estimate the required risk-based profit margin.

Assume that:

1. The probability of zero losses to the catastrophe cover is .96, and the probability that the losses will be \$25 million is .04. Therefore, the variance of the losses is 24 trillion, and the expected losses are \$1 million.
2. Property premium earned for the year is \$100 million, and there is no casualty premium.
3. The standard deviation of pre-tax underwriting return is 15 million.
4. The expected pre-tax return from the entire underwriting portfolio is \$8 million.
5. Taxes have the same proportional effect on the expected pre-tax returns on total premium and on the catastrophe cover, and on the standard deviations of returns.
6. The covariance between the catastrophe cover's losses and total property losses net of the cover is equal to .50 times the variance of the cover's losses.
7. The discount rate for losses is zero.

8. Total underwriting return, and the return on the catastrophe cover, are statistically independent of non-underwriting sources of surplus variability.

It follows from 1, 6 and 8 above, and from the fact that Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z), that the covariance with surplus of the pre-tax return on the catastrophe cover is 24 trillion + .50(24 trillion); i.e., 36 trillion. It follows from 3 and from Cov(X,Y+Z) = Cov(X,Y)+Cov(X,Z) that the corresponding covariance for total underwriting is (15 million)^2, i.e., 225 trillion. Therefore, it follows from assumption 4 that the risk-based profit margin for the catastrophe cover should be such that the pre-tax return from re-assuming the catastrophe cover is given by (36/225)($8 million)= $1.28 million. (This is greater than the cover's expected losses.) If the cover costs more than $2.28 million, then it improves the insurer's risk-return relation to re-assume it. However, the cover may be necessary to maintain the insurer's rating and policyholder comfort.

Required Profit Margin by Layer

Suppose that for some insurer:

1. All premium is property premium.
2. The accident year expected property losses for the $500,000 excess of $500,000 layer, and the 0-$500,000 layer, respectively, are $10 million and $90 million. Expected losses excess of $1 million are zero.
3. The accident year property losses for each of the above layers are independent of all non-underwriting sources of surplus variation.
4. The discount rate is zero.
5. The coefficients of variation (ratios of standard deviations to means) of the higher and lower layers are .30 and .15, respectively.
6. The correlation between the two layers is .5.
7. Taxes have the same proportional effect on the returns of both layers.

Let a and b denote the standard deviations of the losses to the higher and lower layers, respectively. Let ρ denote the correlation. With the above assumptions, the pre-tax covariances with surplus for the higher and lower layers, respectively, are given by

$$
\begin{aligned}
a^2 + \rho ab &= ((10 \text{ million})(.30))^2 \\
&\quad + (.5)(10 \text{ million})(.30)(90 \text{ million})(.15) \\
&= 29.25 \text{ trillion, and} \\
b^2 + \rho ab &= ((90 \text{ million})(.15))^2 \\
&\quad + (.5)(10 \text{ million})(.30)(90 \text{ million})(.15) \\
&= 202.5 \text{ trillion}
\end{aligned}
$$

The allocated surplus for 0-$500,000 layer is 202.5/29.25 (i.e.., 6.9) times as great as the allocated surplus for the $500,000 excess of $500,000 layer. The expected losses are nine times as great for the lower layer. Therefore, the required profit margin, as a percentage of expected losses, is 1.3 (i.e., ((9)(29.25))/202.5) times as great for the higher layer as it is for the lower layer. This is expected due to the higher layer's larger coefficient of variation.

4. RISK-BASED PREMIUM AND THE IRR MODEL

An example is given below to show the following. Suppose that the risk-based premium of my model, for a certain contract corresponds to a certain expected rate of return for the insurer. Then, the expected rate of return for the contract, using the IRR model, also equals that target rate if the method of allocating surplus for the IRR model is the covariance method of my model.

Suppose the target rate of return is 15%. The risk-based premium for a contract equals expense provision + risk-based profit margin + risk-based discounted losses. Suppose that premium and expenses are paid at the end of the year, and the expected loss payout is $100 at the end of each year for four years. Suppose expenses are $70 and the risk-based profit margin is $30. Suppose that the risk-based discount rate is 4%. It then follows that the risk-based discounted losses at the end of the year equal $377.51 and the risk-based premium equals $70 + $30 + $377.51, or $477.51.

By the words surplus and return, I will mean them as defined in my model. The portion of surplus allocated to the contract for the first year will be called S_1 and, using (3.3), it equals

(Surplus)((Covariance(Contract's After-Tax Underwriting Return, Surplus))/(Variance of Surplus))

It is possible to estimate the taxes corresponding to underwriting return at the end of the year of a contract, and the taxes corresponding to the return on risk-based discounted loss reserves and matching assets in the following years. The effect on taxes of premium earned, expenses incurred, investment income from premium, and losses paid during the year, as well as the effect of loss reserves discounted at the beginning and the end of the year, can be used. In the case of risk-based discounted loss reserves and matching assets, the expected taxes are less than taxes on the matching assets. This is because loss reserves are discounted from a point in time one year later at the end of the year than at the beginning of the year, producing a loss for tax purposes. If the discount rate used for tax purposes equals the risk-free rate, but the tax law payout rate is faster than the actual payout rate, the tax effect is the same as the effect of using the actual payout rate and a certain discount rate which is lower than the risk-free rate.

Assume for simplicity that the insurer's assets earn 6% for the period of the coming year and that the tax produced by each type of return is 35% of the return. Then, the above \$30 risk-based profit margin and the above allocated surplus S_1 satisfy the equation

$$.15S_1 = .65(.06S_1+30)$$

since the 15% target return on allocated surplus is produced by the remainder, after 35% tax, of 6% investment income on allocated assets plus the \$30 risk-based profit margin. Solving the equation gives $S_1 = \$175.68$.

The surplus allocated for the second year equals

(Surplus)((Covariance(Contract's After Tax Loss Reserve Return, Surplus))/ (Variance of Surplus))

This allocated surplus will be called S_2. The amount of surplus allocated the next two years are defined similarly and will be called S_3 and S_4.

The risk-based discounted loss reserve corresponding to S_2 equals \$277.51 and satisfies the equation

$$.15S_2 = .65(.06S_2 + (.06 - .04)\$277.51) = .65(.06S_2 + 5.55)$$

since the 15% return is equal to the after-tax return from investment income from the allocated assets plus the after-tax return on loss reserves and matching assets. The matching assets earn 6% and the discounted reserves increase at a rate of 4%, i.e., the risk-based discounted rate. Solving the equation gives S_2 = \$32.50.

Similarly, the risk-based discounted loss reserves corresponding to S_3 and S_4 are \$188.61 and \$96.15, respectively. Therefore,

$$.15S_3 = .65(.06S_3 + (.06 - .04)188.61) = .65(.06S_3 + 3.77)$$
$$.15S_4 = .65(.06S_4 + (.06 - .04)96.15) = .65(.06S_4 + 1.93)$$

So S_3 = \$22.09, and S_4 = \$11.26. It will now be shown that the return on the contract is 15% according to the IRR model, using the same allocation of surplus as above. It was shown above that

$$.15S_1 = .65(.06S_1 + 30) \quad (4.1)$$
$$.15S_2 = .65(.06S_2 + 5.55) \quad (4.2)$$
$$.15S_3 = .65(.06S_3 + 3.77) \quad (4.3)$$
$$.15S_4 = .65(.06S_4 + 1.93) \quad (4.4)$$

If S_1, S_2, S_3, and S_4 are added, respectively, to both sides of the four equations above, respectively, and each equation is divided on both sides by 1.15, we get

$$S_1 = (1/1.15)(S_1 + .65(.06S_1 + 30)) \quad (4.5)$$
$$S_2 = (1/1.15)(S_2 + .65(.06S_2 + 5.55)) \quad (4.6)$$
$$S_3 = (1/1.15)(S_3 + .65(.06S_3 + 3.77)) \quad (4.7)$$
$$S_4 = (1/1.15)(S_4 + .65(.06S_4 + 1.93)) \quad (4.8)$$

Therefore,

$$S_1 = (1/1.15)(S_1 - S_2 + .65(.06S_1 + 30)) + (1/1.15)S_2 \quad (4.9)$$
$$S_2 = (1/1.15)(S_2 - S_3 + .65(.06S_2 + 5.55)) + (1/1.15)S_3 \quad (4.10)$$
$$S_3 = (1/1.15)(S_3 - S_4 + .65(.06S_3 + 3.77)) + (1/1.15)S_4 \quad (4.11)$$

By substituting the expression which is equal to S_4 in equation (4.8) for S_4 in Equation (4.11), we get

$$S_3 = (1/1.15)(S_3 - S_4 + .65(.06S_3 + 3.77)) + (1/1.15)^2 (S_4 + .65(.06S_4 + 1.93))$$

By substituting the above expression for the term S_3 at the extreme right of Equation (4.10), we get

$$\begin{aligned} S_2 = {} & (1/1.15)(S_2 - S_3 + .65(.06S_2 + 5.55)) + \\ & (1/1.15)^2(S_3 - S_4 + .65(.06S_3 + 3.77)) + \\ & (1/1.15)^3 (S_4 + .65(.06S_4 + 1.93)) \end{aligned}$$

By substituting this expression for the term S_2 at the extreme right of Equation (4.9), we get

$$\begin{aligned} S_1 = {} & (1/1.15)(S_1 - S_2 + .65(.06S_1 + 30)) + \\ & (1/1.15)^2(S_2 - S_3 + .65(.06S_2 + 5.55)) + \\ & (1/1.15)^3(S_3 - S_4 + .65(.06S_3 + 3.77)) + \\ & (1/1.15)^4(S_4 + .65(.06S_4 + (1.93)) \end{aligned}$$

Therefore, S_1 is the discounted value, at a 15% return, of amounts at the end of years 1,2,3 and 4 which are each equal to the following: the sum of supporting surplus which is no longer needed at the end of the year plus the after-tax return during the year resulting from investment income from supporting surplus and from the contract. So, according to the IRR model, the rate of return on the contract is 15%. This completes the demonstration of the relationship between risk-based premium and the IRR model.

REFERENCES

[1] Gogol, Daniel F. "Pricing to Optimize an Insurer's Risk-Return Relation," PCAS LXXXIII, 1996, pp.41-74.

[2] Feldblum, Sholom, "Pricing Insurance Polices: The Internal Rate of Return Model," CAS Study Note, May 1992.

[3] Bingham, Russell E., "Surplus – Concepts, Measures of Return, and Determination," PCAS LXXX, 1993, pp.55-109.

[4] Bingham, Russell E., "Policyholder, Company, and Shareholder Perspectives," PCAS LXXX, 1993, pp.110 – 147.

[5] Markowitz, Harry, "Portfolio Selection," The Journal of Finance, March 1952, pp. 77-91.

Fitting to Loss Distributions with Emphasis on Rating Variables

Farrokh Guiahi, Ph.D., FCAS, ASA

Abstract

This paper focuses on issues and methodologies for fitting alternative statistical models--probability distributions--to samples of insurance loss data. The interaction of parametric loss distributions, deductibles, policy limits and rating variables in the context of fitting distributions to losses are discussed. Fitted loss distributions serve an important function for the purpose of pricing insurance products. The procedures illustrated in this paper are based on a sample of insurance losses, and with lognormal as the underlying loss distribution.

Key words

Loss Distributions, Generalized Linear Models, Curve Fitting, Right Censored and Left Truncated data, Rating Variables, Maximum Likelihood Estimation.

1. Introduction

This section presents some preliminaries regarding losses, deductibles, policy limits and rating variables as inputs for fitting distributions to losses. In section 2, a method for fitting a single distribution to losses is considered. In this instance, the information provided by rating variables is either not considered or is not available. The method of maximum likelihood has been applied to estimate model parameters in the presence of deductibles and policy limits. Sections 3 and 4 develop methodologies for fitting alternative statistical models--family of loss distributions--to loss data, using the information provided by rating variables. This is achieved by requiring a parameter of a loss distribution to depend upon values of rating variables. Criteria for assessing goodness of fit are discussed. Furthermore, large sample statistical tests for assessing the impact of rating variables upon loss distributions are given. Some concluding statements are made in section 5.

Insurance data considered here have the following characteristics: a) losses are specified individually, b) for each individual loss, the information about deductibles and policy limits is furnished, and c) for each loss, we have auxiliary policy information regarding the rating variables. Each of these three items is discussed further below.

Losses are given on an individual basis, and have not been grouped by loss size. The methodologies to fit distributions to data differs, depending on whether losses are grouped or individually specified. Losses may be closed or open. The amount recorded for each loss is the incurred value as of the latest available evaluation period. If some losses in the sample data are still open as of the latest evaluation period, then those losses should be properly adjusted for further development. Unfortunately, most of the methodologies for development of losses to their ultimate values are only available for grouped data. Further research on the topic of development of individual losses to their individual ultimate values is welcomed. Individual losses should be suitably trended to reflect values expected in the future. The methodology presented in this paper has been applied to a sample of commercial fire losses (see Table A of Appendix A). Those losses were mostly closed, as of their latest evaluation date, hence adjustments for further development were not warranted. Finally, in order to fit distributions to losses, zero losses should be excluded.

Deductibles are used to exclude certain losses. Usually deductibles are small--for example, a few hundred or a few thousand dollars. However, for a large insured, deductibles may be sizable due to the existence of self-insured retention or other underlying coverages. Only dollar deductibles are considered here. Time deductibles such as waiting periods are not treated. A reported loss with a value in excess of its deductible is said to be left truncated. If a loss arises from a policy with no underlying deductible, then for the purpose of the computation, a value of zero is imputed as the "deductible" amount. It is not required that the deductible amount be the same for each loss.

Policy limits serve to limit the amount of payment on a given loss or a loss occurrence. When the loss amount is at least as large as its policy limit, the loss is said to have been right censored. If a loss arises from a policy where there is no underlying policy limit, then any amount greater than the loss amount may be imputed as the "policy limit". In these instances, those losses have not been censored. Varying policy limits are allowed for. In fact, no grouping of losses based upon deductible or policy limit amounts is required.

Samples of insurance loss data are usually incomplete. This is due to inclusion of left truncated (losses in excess of deductibles) and right censored (some losses capped by their respective policy limits) data in the sample. Due to this incompleteness of data, it becomes more difficult to estimate the parameters of a loss distribution and to assess the goodness of fit. Many traditional approaches for estimation of parameters of a loss distribution or assessing the goodness of fit of a distribution are valid only if the sample of observations is complete, that is, when there are neither left truncated nor right censored observations in the sample.

Rating variables in insurance depend upon the line of business, the degree of competition present in the market, and regulation. The effect of the rating variables upon loss distributions has important implications for underwriting selection. It also provides for a more differentiated rating system. How to incorporate the information provided by rating variables into the process of fitting distributions to losses is discussed in sections 3 and 4.

Following is a description of how to fit a single distribution to a sample of insurance loss data.

2. Fitting a Single Distribution to Losses

Fitting a single distribution to losses is based upon consideration of alternative statistical models--probability distributions--as data-generating mechanisms. The assumption made is that the observed losses are a realization of a probabilistic process governed by a parametric distribution. The purpose of fitting a distribution to losses is to identify a specific parametric distribution which provides a reasonable fit to the data. A good introduction to the subject of fitting distributions to losses is given by Hogg and Klugman (1984). This paper complements their work by focusing on certain related topics. First, more emphasis is placed on the procedures for fitting loss distributions to individual loss data rather than grouped data. Second, methodologies required to incorporate rating variables in the process of fitting distributions to losses are presented in sections 3 and 4. Finally, readers of this paper may find the computer programs (codes) given here to be beneficial for the purpose of the computing maximum likelihood estimates of parameters of a loss distribution.

Fitting a distribution to losses serves to moderate the effect of sampling variation in the data. This is achieved by replacing an empirical distribution by a more smoothed (fitted) distribution. Furthermore, estimates of tail probabilities beyond the range of the original data can be provided based on fitted distribution.

At least two problems complicate the fitting of a parametric distribution to loss data. The first problem concerns the tendency of many losses to be settled at rounded figures. This notion is incompatible with selecting a parametric distribution such as lognormal or Pareto, where the probability of taking any specified value is zero. The second problem arises from the fact that many statistical procedures assume that losses in a sample are identically distributed. Insurance risks are normally heterogeneous. Each risk has its own risk characteristics and its own propensity to produce a potential loss. For instance, two different drivers have differing loss propensities. To a certain extent, risk characteristics are reflected by underwriting rating factors. For this reason, risks with the same values for their underwriting factors are cross-classified to produce "homogeneous" classes. The use of rating factors to cope with the heterogeneity problem is addressed in sections 3 and 4. In this section, the information provided by rating factors is ignored in order to concentrate on fitting a single loss distribution to data.

For the sake of exposition, the process of fitting a single distribution to loss data has been broken down into four steps:

1. Consideration of a number of parametric probability distributions as potential candidates for underlying loss distribution.
2. For each distribution specified in step 1, the estimation of the parameters of the distribution from sample data--hence, the determination of a set of fitted distributions.
3. Specification of a criterion for choosing one or a few fitted distributions from step 2 above.
4. Assessing the goodness of fit for the fitted distribution(s) in step 3.

Let us proceed with a more detailed account of these steps. These steps will be illustrated below by reference to a numerical example. The first step requires considering a number of parametric distributions as potential candidates for the data generating mechanism. The list of potential parametric distributions as candidates for loss distribution is enormous. In practice, one can entertain only a few parametric distributions for the purpose of fitting a distribution to losses. In this paper, I have selected the following parametric probability distributions: lognormal, Pareto, Weibull, gamma, inverse gamma, and exponential. This list is subjective, but some of the above distributions have been used by actuaries and have appeared in actuarial literature. The list chosen here is only for illustrative purposes and is not meant to be exhaustive.

The second step involves the estimation of the parameters of each probability distribution selected in step 1 from the data. Once one has estimated the parameters of a given distribution, one then has a fitted distribution. The estimation of parameters of a loss distribution is made difficult because of incompleteness of data. Some commonly used statistical procedures to estimate parameters of a distribution for a sample of complete data are: the method of moments, the least squares estimation as used for regression models, and the maximum likelihood estimation. These parameter estimation procedures are outlined in most basic statistics texts. For incomplete sample data (presence of left truncated or right censored data), the above estimation procedures are not applicable without further modifications. The application of estimation procedures suitable for complete data to insurance data which is incomplete will produce inefficient parameter estimates. In this paper, the estimation of parameters of a loss distribution is based upon proper specification of the likelihood function reflecting the presence of left truncated and right censored observations in the data.

Following are some necessary notations needed to write an expression for the likelihood function in the case of incomplete data.

Let y_i be the i^{th} loss amount (incurred value), $1 \leq i \leq n$, where n denotes the number of losses in the data set.

D_i is the deductible for the i^{th} loss.

PL_i is the policy limit for the i^{th} loss.

$f(y;\theta,\varphi)$ denotes the density function for the loss amount in the case of complete data. θ is the primary parameter of interest. φ is the nuisance parameter which may be a vector.

$F(y;\theta,\varphi)$ denotes the cumulative distribution function for the loss amount.

The contribution of a loss to the functional form of the likelihood function depends upon whether the loss is ground-up or in excess of deductible, and furthermore if the loss has been capped by its respective policy limit. Hence, the contribution of a loss to the likelihood function may be one of the four mutually exclusive and exhaustive forms, written as L_{i1}, L_{i2}, L_{i3}, and L_{i4}, as defined below. In addition, four indicator variables, δ_{i1}, δ_{i2}, δ_{i3} and δ_{i4} are used in order to write a succinct expression for the likelihood function of the sample.

Case 1: No deductible, and loss below policy limit (neither left truncated nor right censored data). The complete sample case.

$$L_{i1} = f(y_i;\theta,\varphi) \qquad (2.1a)$$

$$\delta_{i1} = \begin{cases} 1, \text{If } D_i = 0 \text{ and } y_i < PL_i \\ 0, \text{ Otherwise} \end{cases} \qquad (2.1b)$$

Case 2: A deductible, and loss below policy limit (left truncated data)

$$L_{i2} = \frac{f(D_i + y_i;\theta,\varphi)}{1 - F(D_i;\theta,\varphi)} \qquad (2.2a)$$

$$\delta_{i2} = \begin{cases} 1, \text{If } D_i > 0 \text{ and } y_i < PL_i \\ 0, \text{ Otherwise} \end{cases} \qquad (2.2b)$$

Case 3: No deductible, and loss capped by policy limit (right censored data)

$$L_{i3} = 1 - F(PL_i;\theta,\varphi) \qquad (2.3a)$$

$$\delta_{i3} = \begin{cases} 1, \text{If } D_i = 0 \text{ and } y_i \geq PL_i \\ 0, \text{ Otherwise} \end{cases} \qquad (2.3b)$$

Case 4: A deductible, and loss capped by policy limit (left truncated and right censored data)

$$L_{i4} = \frac{1 - F(D_i + PL_i;\theta,\varphi)}{1 - F(D_i;\theta,\varphi)} \qquad (2.4a)$$

$$\delta_{i4} = \begin{cases} 1, \text{If } D_i > 0 \text{ and } y_i \geq PL_i \\ 0, \text{ Otherwise} \end{cases} \qquad (2.4b)$$

The contribution of the i^{th} loss to the likelihood function is given by

$$L_i = L_{i1}^{\delta_{i1}} L_{i2}^{\delta_{i2}} L_{i3}^{\delta_{i3}} L_{i4}^{\delta_{i4}} \qquad (2.5)$$

The likelihood function for the sample is given by

$$L = \prod_i L_i \qquad (2.6)$$

The log-likelihood is given by

$$l = \sum_i \log(L_i) \qquad (2.7a)$$

$$= \sum_i l_{it} \qquad (2.7b)$$

$$l_i = \log(L_i) \qquad (2.8a)$$

$$= \delta_{i1} \log(L_{i1}) + \delta_{i2} \log(L_{i2}) + \delta_{i3} \log(L_{i3}) + \delta_{i4} \log(L_{i4}) \qquad (2.8b)$$

where the log, as used in this paper, represents the natural logarithm.

Equation (2.5) represent the contribution of the i^{th} loss to the likelihood function. The likelihood function for the data is given by equation (2.6). To estimate the parameters θ and φ we should maximize the likelihood function or alternatively minimize the negative of the logarithm of the likelihood function.

Equation (2.7) and (2.8) provide expressions for the logarithm of the likelihood function.

Note that the contribution to the likelihood function for an individual observation in most basic statistics textbooks is of the form (2.1a).

The third step requires a criterion for ranking or comparing alternative fitted probability distributions. This step is needed to reduce the number of fitted distributions in step 2 to one or a few potential candidates. A statistical criterion used for comparing alternative models--statistical distributions--is based upon the value of Akaike's Information Criterion, AIC; refer to Akaike (1973).

The AIC criterion is defined by

AIC = - 2(maximized log-likelihood)
+ 2(number of parameters estimated)

Note, AIC can also be written as

AIC = - 2{maximized log-likelihood - number of parameters estimated}

When two models are compared, the model with a smaller AIC value is the more desirable one.

The AIC is based on log-likelihood and it penalizes the log-likelihood by subtracting for the number of parameters estimated. Two other model selection criteria used in statistics are Schwarz's Bayesian Information Criterion (BIC), Schwarz(1978), and Deviance as used in Generalized Linear Models; see McCullagh and Nelder (1989). These three criteria are based on maximized log-likelihood function.

Before proceeding to step 4, regarding fit, I shall illustrate steps 1, 2, and 3 by reference to a numerical example. Let us consider the data in Table A of Appendix A. Here, we have a sample of 100 commercial fire losses. For each loss the deductible, policy limit, and the code for a type of construction are stated. For the time being, let us ignore the information about the construction since we are concerned with fitting a single distribution to the data. For each distribution listed in Table 1 below, I have computed the maximized log-likelihood function, and the corresponding AIC values. For the case of Weibull distribution, the program used to compute the maximum likelihood estimate of parameters and the computed value of maximized log-likelihood function is given as Exhibit 1 in Appendix B. This program is coded in S-Plus, a statistical software suitable for data analysis. The computation of maximized likelihood function for other distributions in Table 1 is similar to the one for Weibull.

Table 1

Distribution	Negative maximized log-likelihood function	Number of Parameters	AIC
lognormal	897.8	2	1799.6
Pareto	895.2	2	1794.4
Weibull	899.8	2	1803.6
gamma	914.5	2	1833.0
inverse gamma	893.7	2	1791.4
exponential	986.4	1	1974.8

With regard to Table 1, it should be noted that the values of maximized likelihood function are positive. The values of logarithm of the maximized likelihood functions are negative and hence the negatives of the logarithm of the maximized likelihood functions are positive figures.

Table 1 can be used for selecting a parametric distribution for the data. Based on the AIC criterion as a method of ranking different fitted distributions, note that the AIC values of lognormal, Pareto, and inverse gamma are "comparable". The AIC values for gamma and exponential distributions suggest relatively more inferior fits. I have selected lognormal, with parameters μ and σ^2, as the distribution to be fitted to our data. There are several reasons for this selection. First, it is easier to interpret the parameters of a lognormal distribution. Selecting a simpler model is preferable, as it is easier to explain and comprehend. By taking the logarithm of the losses, the μ parameter represents the location parameter (mean), and the σ parameter is the scale (standard deviation). Second, lognormal distribution has been previously used to describe the distribution of fire losses; see Benckert and Jung (1974).

Now we proceed with step 4, regarding the fit. By examining the data in Appendix A, we note that the losses can be divided into four categories according to four cases defined for specification of the likelihood function (see Table 2 below):

Table 2

#	Case	Number of Losses
1.	No deductible and loss below policy limit	1
2.	A deductible, and loss below policy limit	96
3.	No deductible and loss capped by policy limit	0
4.	A deductible and loss capped by policy limit	3

For our data, most of the losses are of case 2, i.e., losses with deductibles and values below their policy limits. Due to the paucity of data, we concentrate only on case 2.

For lognormal distribution, we can compute theoretical conditional distributions (probabilities) and conditional limited expected values based on a fitted distribution, and compares these quantities with their respective sample counterparts.
The conditional distribution or probability of a lognormal random variable, X, with parameters μ and σ is given by

$$P(X \le b \mid X > a) = \frac{\Phi(\frac{\log(b)-\mu}{\sigma}) - \Phi(\frac{\log(a)-\mu}{\sigma})}{1-\Phi(\frac{\log(a)-\mu}{\sigma})}$$

where Φ is the cumulative distribution function of a standard normal distribution. Here a represents a threshold or a deductible amount D, and b is usually the sum of deductible and limit, i.e., D + PL.

The conditional limited expected value is defined by

$$E[\min(X,b) \mid X > a] = \frac{1}{1-\Phi(\frac{\log(a)-\mu}{\sigma})}\{ e^{\mu+\frac{1}{2}\sigma^2} [\Phi(\frac{\log(b)-\mu-\sigma^2}{\sigma}) - \Phi(\frac{\log(a)-\mu-\sigma^2}{\sigma})] + b[1-\Phi(\frac{\log(b)-}{\sigma}$$

Table 3 summarizes the comparison of theoretical and sample values of conditional probabilities and conditional limited expected values for case 2 of data in Appendix A.

Table 3

Comparison of Conditional Probabilities and Conditional Limited Expected Value for Fitted Lognormal with its Sample Values

a = 500

	$P(X \le b \mid X > a)$		$E[\min(X,b) \mid X > a]$	
b	Based on lognormal*	Sample estimate	Based on lognormal*	Sample estimate
2,000	0.485	0.494	1,538.7	1,620.9
5,000	0.714	0.699	2,666.4	2,737.2
10,000	0.832	0.843	3,747.2	3,764.3
20,000	0.909	0.904	4,969.3	4,907.7
30,000	0.938	0.952	5,716.8	5,547.9
40,000	0.954	0.976	6,248.3	5,833.6
50,000	0.964	0.988	6,655.8	6,071.7

* $\hat{\mu}$ = 5.887, $\hat{\sigma}$ = 2.302 are the maximum likelihood estimates for the fitted lognormal distribution.

The comparisons of fitted and sample quantities in Table 3 suggests the lognormal provided a "reasonable" fit to the data. It is worth making a few comments regarding fit. First, our sample size is 100, with 96 observations for case 2. With small sample sizes, considerable sampling variability are encountered in estimation of model parameters. Second, a perfect fit implies no smoothing! Third, the fit for a specific type of distribution is judged to be good if it has a high predictive power, that is, whether the same type of distribution provides good fits to many samples of the same kind. A quotation from Lindsey (1995), is appropriate here: "*If a model represents the sample too well, it will have no chance of representing a second, similarly generated, sample very well. A model too close to a sample will usually be too far from the population.*" Finally, it is worth emphasizing that there are many other possible potential candidates (probability distributions) for fitting to a specific data set. Thus, curve fitting is to some extent subjective and not a perfect science. From a practical point of view, there are other considerations related to fitting a distribution to a sample. These are: a) the volume and quality of data, b) the time constraint in which to do the curve fitting, c) the knowledge and experience of the curve fitter, d) availability of suitable software (programs), e) convergence of iterative algorithms for estimation of model parameters, and specification of initial values for parameters, and f) the treatment of outliers. Last but not the least is consideration of the purpose for which the fitted distribution is used. With all these qualifications regarding fit, we shall assume the lognormal provides a reasonable fit to the data in Appendix A.

3. Fitting a Family of Distributions to Loss Data: A Mean Approach

In section 2, procedures to fit a single distribution to loss data were considered. The information provided by rating variables was not considered. As mentioned earlier, risks in insurance tend to be heterogeneous. Risks with different attributes may well have different loss distributions. To a certain degree, a risk's characteristics are reflected through the values pertained by its rating variables. Thus, we expect the loss distribution for fire for a small unprotected frame building be different from a large, highly protected and fire-resistive building. It is desirable to have loss distributions which reflect these differences. Our approach to this issue is to construct suitable statistical models--family of loss distributions. Two possible solutions are proposed in this paper. The first solution, as explained in this section, is similar in spirit to the Generalized Linear Models (GLM) approach. An excellent account on the subject of GLM is given by McCullagh and Nelder (1989). An alternative solution is presented in section 4.

Loss distributions dependent upon rating variables have important implications for underwriting selection and determination of rates. By including the rating variables, one generally improves the fit to the data. Using statistical models enables one to assess the effect of rating variables on loss distributions by performing statistical tests of hypotheses.

A traditional approach for obtaining loss distributions dependent upon risk attributes is to segment losses into subgroups. Then, for each subgroup, a separate fitted loss distribution is obtained. For instance, in fire insurance, losses may be classified broadly by construction as frame, masonry and fire-resistive. Three fitted loss distributions can be obtained according to the types of construction. Segmentation of data into classes gives rise to credibility problems. For the problem alluded to, it would be exasperating if one considered eight construction types instead of three, and in addition, considered other rating factors such as protection and occupancy.

In section 2, we noted that the lognormal distribution provides a reasonable fit to the data in Appendix A. Mirroring the approached used in GLM, let us now fit a family of lognormal distributions to our data.

The GLM methodologies consist of three components. These are referred to as the <u>random component</u>, the <u>systematic component</u>, and the link. The random component: the random variable of interest, Y (e.g., losses) or a transformation of Y, has a distribution belonging to the exponential family of distributions. The density, in canonical form, for the exponential family is

$$f(y;\theta,\varphi) = \exp\{[(\theta y - b(\theta)) / a(\varphi)] + c(y,\varphi)\}$$

where a(.), b(.) and c(.) are some specific functions. θ is the primary parameter of interest, and φ is often referred to as the nuisance parameter. Suitable loss distributions in the exponential family include normal, gamma and inverse Gaussian.

The systematic component of a GLM specifies the explanatory variables, $x_1, x_2, \ldots, x_p$ (e.g., rating variables). The explanatory variables may only influence the distribution of the Y through a single linear function called the linear predictor η,

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

The link, g, specifies how the mean of Y, E(Y), is related to the linear predictor η, i.e.

$$g(E(Y)) = \eta = \sum_j \beta_j x_j$$

The form of the link function varies by the type of distribution within the exponential family of distributions. For the normal distribution the link function is the identity map, i.e., $\mu = \eta$.

In GLM, the information provided by explanatory variables (rating variables) is summarized by a linear predictor.

Each explanatory variable is considered either as a factor (categorical) or as a covariate (quantitative). For instance, sex, construction, and protection are categorical in nature, while age and amount of insurance are quantitative.

Some additional notations are needed to specify our statistical model. Let η_i denote a linear predictor for the i^{th} loss. It summarizes the information conveyed by the rating variables for the i^{th} loss. We write

$$\begin{aligned}\eta_i &= x_i^T\beta \\ &= \sum_{j=0}^{p} x_{ij}\beta_j \\ &= \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\end{aligned}$$

where β is a $(p+1)\times 1$ vector of unknown parameters. x_i is a $(p+1)\times 1$ vector of known constant terms, x_{ij}'s. The first element of x_i, x_{i0} is set equal to one. Its purpose is to represent a constant term (intercept) in the expression for the linear predictor. The other x_{ij}'s components, $1 \le j \le p$, are used to represent rating variables. The value of p is partially dependent upon the number of categorical rating factors included in the model, as well as their respective number of levels (values). In addition, p depends upon the number of quantitative rating variables in the model. Note that when rating variables are not taken into consideration, or the information about them is not available, then p takes on the value of zero. This corresponds to the fitting of a single distribution to the entire loss data as described in section 2.

Following are some examples of the linear predictors, η_i, to be discussed throughout this paper. Some commonly used categorical rating factors in fire insurance are construction, protection, and occupancy. The amount of insurance (insured building value), a measure of exposure, is quantitative. Here, we shall consider only construction and building value for illustrative purposes. Assume there are three possible construction types (levels), namely frame, masonry and fire-resistive. In GLM, as well as regression analysis, the contribution of a categorical variable to a linear predictor is made by specifying dummy variables. For the construction rating factor, we need to introduce two dummy variables C_{i1} and C_{i2}, defined as follows:

$$C_{i1} = \begin{cases} 1, \text{If the i}^{\text{th}} \text{ risk is a frame} \\ 0, \text{Otherwise} \end{cases}$$

$$C_{i2} = \begin{cases} 1, \text{If the i}^{\text{th}} \text{ risk is a masonry} \\ 0, \text{Otherwise} \end{cases}$$

For the i^{th} loss, let BV_i denote the amount of insurance purchased by the policyholder to cover damages arising from peril of fire to the building. For a fire policy, the policy limit for the building cover is synonymous with the building value. Since there is a wide range of variability among building values, we shall use the logarithm of the building value instead of building value as our covariate in the linear predictor. For these two variables, namely, construction and building value, we shall define four statistical models corresponding to four linear predictors as follows:

Model A: $\eta_i = \beta_0$ (3.1A)

Model B: $\eta_i = \beta_0 + \beta_1 C_{i1} + \beta_2 C_{i2}$ (3.1B)

Model C: $\eta_i = \beta_0 + \beta_1 \log(BV_i)$ (3.1C)

Model D: $\eta_i = \beta_0 + \beta_1 \log(BV_i) + \beta_2 C_{i1} + \beta_3 C_{i2}$ (3.1D)

The linear predictor given by equation (3.1A) is used when either we do not take into consideration the information given by rating variables or when no information on rating variables is available. In these instances, we are fitting a single distribution to the entire data. We shall refer to this Model A as the "base" model (distribution). The base distribution is used as a benchmark to gauge the relative improvement in fit by including rating variables.

The linear predictor corresponding to (3.1B) is appropriate if construction is the only rating factor used. Using the statistical methodology developed here, the entire data is used to estimate the values of the parameters β_0, β_1, β_2 simultaneously. This approach is different from the one in which the data is segmented into three sub-groups according to types of construction.

The linear predictor (3.1C) is used when we wish to examine only the effect of exposure size (building value) on loss distribution.

Finally, we shall use (3.1D) when both construction and building value are considered. In this case, the vector $x_i^T = (1 \;\; \log(\mathrm{BV}_i) \;\; C_{i1} \;\; C_{i2})$ represents the contribution of the i^{th} risk's attributes to the linear predictor, and p has the value of three.

The four linear predictors given by (3.1A), (3.1B), (3.1C), and (3.1D) generate four statistical models. This is an example of nested models. For nested models, some models are a special case of a more general model. The linear predictors (3.1A), (3.1B) and (3.1C) are special cases of the linear predictor (3.1D). For the linear predictor (3.1D), Model D, we can entertain the following statistical tests of hypotheses:

$$H_0\colon \beta_1 = \beta_2 = \beta_3 = 0 \qquad (3.2)$$

$$H_0\colon \beta_2 = \beta_3 = 0 \qquad (3.3)$$

$$H_0\colon \beta_1 = 0 \qquad (3.4)$$

The null hypothesis (3.2) is used to test if either construction or building value (exposure size) has any effect on loss distribution. The acceptance of this null hypothesis, subject to the usual interpretation of Type Two error probability, suggests that the rating variables have no appreciable influence on the loss distribution. The rejection of (3.2) implies that the inclusion of building value or construction in the linear predictor gives a superior model as compared to the fit by the base distribution, Model A. The acceptance of the null hypothesis (3.3) suggests that in the presence of building value, the addition of the construction factor does not improve the fit. Null hypothesis (3.4) can be similarly interpreted.

By conducting statistical tests corresponding to the previously stated hypotheses, the effects of rating variables on loss distributions can be assessed. The test statistics are likelihood ratio tests. The asymptotic distribution of test statistics are Chi-squares. Hence, for small sample sizes, the implications of the above tests based on Chi-squares are only approximately valid.

Here, we assume that the underlying loss random variable, Y_i--for the i^{th} risk--has a lognormal distribution with parameters μ_i and σ^2. The parameter μ_i is the mean of transformed variable $\log(Y_i)$. We shall refer to models in this section as "Mean" models. Using an approach similar to GLM, we relate the rating variables of interest to parameter μ_i by using an identity link function. That is,

$$\begin{aligned}\mu_i &= x_i^T\beta \\ &= \beta_0 + \sum_i x_{ij}\beta_j\end{aligned}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are regression like parameters and x_{ij}'s represent the contribution of explanatory rating variables for the ith risk. Hence, we have a family of lognormal distributions, with parameters $\beta_0, \beta_1, \ldots, \beta_p$ and σ^2 to describe the distribution of losses.

It is assumed that the parameter σ is the same for each risk, and does not vary by the rating variables. We shall examine an alternative approach in the next section, where σ is not constant. Although, the mean and variance of the loss distributions vary by rating variables, but due to the constancy of σ, the skewness, and the kurtosis are not dependent on rating variables.

The mechanism to fit a family of lognormal distributions to the data of Table A of Appendix A has now been established. A set of nested hypotheses of interest, (3.2), (3.3), and (3.4) in reference to model (3.1D) has also been stated. We now need to perform the necessary computations to estimate the model parameters, and calculate log-likelihood statistics for alternative models as described by linear predictors (3.1A), (3.1B), (3.1C), and (3.1D).

The program to compute maximum likelihood estimate of model parameters for the linear predictor (3.1D), as well as the value of the <u>negative</u> of log-likelihood based upon maximum likelihood estimates is given as Exhibit 2 of Appendix B.

Likelihood ratio test statistics are needed for performing nested statistical tests of hypothesis (3.2), (3.3), and (3.4). The likelihood ratio test statistics can be calculated from the values of log-likelihood statistics for the appropriate models.

The upper portion of Table 4 below provides the values of the negative of log-likelihood statistics for the " mean" models according to linear predictors (3.1A), (3.1B), (3.1C), and (3.1D). The lower portion of Table 4, provides the values of the necessary likelihood ratio test statistics for performing nested statistical hypotheses (3.2), (3,3), and (3.4). In addition, the appropriate 95^{th} percentiles and degrees of freedom of the asymptotic distributions of test statistics are also provided.

Table 4

Likelihood Statistics for Alternative Statistical Models

"Mean" Models

Model	Linear Predictor	Negative of logarithm of Likelihood function
A	$\mu_i = \beta_0$	897.7654
B	$\mu_i = \beta_0 + \beta_1 C_{i1} + \beta_2 C_{i2}$	894.8344
C	$\mu_i = \beta_0 + \beta_1 \log(BV_i)$	896.8284
D	$\mu_i = \beta_0 + \beta_1 \log(BV_i) + \beta_2 C_{i1} + \beta_3 C_{i2}$	892.7099

Nested Hypotheses based on Model D

Test of Hypothesis	Likelihood Ratio* Test Statistics	DF for Chi-sq.	95^{th} perc. of Chi-sq.
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$	$-2(\log L_A - \log L_D) = 10.1110$	3	7.8147
H_0: $\beta_2 = \beta_3 = 0$	$-2(\log L_C - \log L_D) = 8.2370$	2	5.9915
H_0: $\beta_1 = 0$	$-2(\log L_B - \log L_D) = 4.2490$	1	3.8415

* L_A, L_B, L_C, and L_D, above, correspond to likelihood statistics for "Mean" Models A, B, C, and D respectively.

Let us interpret the results given by Table 4, later on we shall make some qualifications regarding our interpretations.

If we are interested to test whether construction factor or building value has an effect upon the shape of the loss distribution, the appropriate null hypothesis is H_0: $\beta_1 = \beta_2 = \beta_3 = 0$. The value of the test statistic, i.e., the likelihood ratio test statistics is 10.111. Since 10.111 exceeds the value of 7.8147 (the boundary of rejection region), it implies that we should reject the null hypothesis H_0. The implication is either construction or building value have an influence on the shape of the loss distribution. Similar interpretations can be given for the other two null hypotheses.

Some qualifications regarding the above interpretation of Table 4 are in order. First, due to relatively small sample size, and the approximate distribution of likelihood ratio test, as Chi-squares, we should be careful to interpret the results given in Table 4. Second, the numerical estimate of parameters (see Exhibit 2 of Appendix B) and the implications of the nested test of hypotheses, are <u>only</u> for illustrative purposes and are <u>not</u> intended to be used for any rating purposes.

Finally, the Model D has the largest likelihood value. Based upon the values of likelihood statistics, as well as the AIC values, Model D fits the data better than Model A, the base distribution. Recall that Model A corresponds to the case of fitting a single distribution to the data. Thus, the consideration of rating variables has led to an improvement in fit, and this improvement is statistically significant.

4. Fitting a family of Lognormal Distributions with Different Scale Parameters

In section 3, a family of lognormal distributions using a procedure " similar" to the GLM approach was introduced. These alternative statistical models were referenced to as "Mean" models. The linear predictor was set equal the μ parameter of the lognormal, and the σ parameter was assumed to be constant. By considering the logarithm of losses, log(Y), the rating variables affected the mean of the distribution but not the scale, the σ parameter. In this section, a family of lognormal distributions is introduced where the scale σ is made to depend on rating variables, and the parameter μ is treated as a constant. Using methodology similar to that in section 3, four new statistical models A, B, C, and D, are defined corresponding to four linear predictors as follows:

Model A: $\sigma_i = \beta_0$ (4.1A)

Model B: $\sigma_i = \beta_0 + \beta_1 C_{i1} + \beta_2 C_{i2}$ (4.1B)

Model C: $\sigma_i = \beta_0 + \beta_1 \log(\mathrm{BV}_i)$ (4.1C)

Model D: $\sigma_i = \beta_0 + \beta_1 \log(\mathrm{BV}_i) + \beta_2 C_{i1} + \beta_3 C_{i2}$ (4.1D)

These models will be referred to as "Scale" models. Parallel to the development in section 3, we have three nested statistical hypotheses of interest for Model D, linear predictor (4.1D), defined as

$$H_0\colon \beta_1 = \beta_2 = \beta_3 = 0 \qquad (4.2)$$

$$H_0\colon \beta_2 = \beta_3 = 0 \qquad (4.3)$$

$$H_0\colon \beta_1 = 0 \qquad (4.4)$$

The purpose and interpretation of these hypotheses is similar to those of (3.2), (3.3), and (3.4) of section 3.

With the mechanism established in section 3, we want to evaluate the it of alternative "Scale" models fitted to the data in Table A of Appendix A. The results of these computations are summarized in Table 5 below. A program for the maximum likelihood estimate of parameters, and likelihood statistics for Model B, linear predictor (4.1B), is given in Exhibit 3 of Appendix B. For comparison purposes, the values of likelihood ratio statistics for the "Mean" models are also reproduced in Table 5.

Table 5

Likelihood Statistics for Alternative Statistical Models

"Scale" Models

Model	Linear Predictor	Negative of logarithm of Likelihood function
A	$\sigma_i = \beta_0$	897.7654
B	$\sigma_i = \beta_0 + \beta_1 C_{i1} + \beta_2 C_{i2}$	892.4242
C	$\sigma_i = \beta_0 + \beta_1 \log(\mathrm{BV_i})$	895.7967
D	$\sigma_i = \beta_0 + \beta_1 \log(\mathrm{BV_i}) + \beta_2 C_{i1} + \beta_3 C_{i2}$	887.9109

Nested Hypotheses Based On Model D Comparison of "Mean" & "Scale" Models

Test of Hypothesis	Likelihood Ratio Test Statistics*	Mean Model	Scale Model	DF for Chi-sq.	95^{th} perc. of Chi-sq.
$H_0: \beta_1 = \beta_2 = \beta_3 = 0$	$-2(\log L_A - \log L_D)$	10.1110	19.7090	3	7.8147
$H_0: \beta_2 = \beta_3 = 0$	$-2(\log L_C - \log L_D)$	8.2370	15.7716	2	5.9915
$H_0: \beta_1 = 0$	$-2(\log L_B - \log L_D)$	4.2490	9.0266	1	3.8415

*Depending upon the context, the L_A, L_B, L_C, and L_D, above, correspond to likelihood functions for "Mean" or "Scale" Models A, B, C, and D.

Once again we should be careful to interpret the results given in Table 5 due to relatively small sample size, and the approximate distribution of likelihood ratio test as Chi-squares. With these qualifications in mind, it appears that the "Scale" models provide a better fit than the "Mean" models to our data.

5. Conclusion

This paper discusses issues related to curve fitting. It provides appropriate statistical methodologies for fitting parametric distributions to loss data. In particular, the interaction of parametric probability distributions, deductibles, policy limits and rating variables are considered. The presence of deductibles and policy limits complicate the estimation of parameters of loss distribution, and the assessment of goodness of fit. Procedures to fit a single distribution or a family of distributions to loss data were given. Statistical tests of hypotheses to assess the effect of rating variables upon loss distribution were discussed. The methodologies developed in this paper were applied to a sample of loss data using lognormal as the reference distribution. Sample programs coded in S-Plus, a statistical package, were provided to illustrate the numerical computation of maximum likelihood estimate of model parameters and maximized likelihood function. Finally, the results in this paper suggest that for any specific data set, there may be many viable statistical models suitable for the purpose of fitting distributions to the data.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csaki, F., *Second International Symposium on Inference Theory*, Budapest: Akademiai Kiado, pp. 267-281.

Benckert, L.G. and Jung, J. (1974). Statistical Models of Claim Distributions in Fire Insurance. *ASTIN Bulletin 8, 1-25.*

Hogg, R.V. and Klugman, S.A. (1984). *Loss Distributions*. John Wiley & Sons, New York.

Lindsey, J. K. (1995). *Introductory Statistics: A Modelling Approach*. Oxford University Press.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Models*, Second Edition. Chapman and Hall, New York.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461-464.

Appendix A: TABLE A

Deductible	Policy Limit	Loss	Construction	Deductible	Policy Limit	Loss	Construction
1,000	57,000	502	2	250	43,000	75	2
250	41,000	31,971	1	1,000	1,000	865	3
1,000	1,000	367	1	100	33,000	206	2
250	60,000	698	2	250	7,000	2,303	1
100	10,000	4,863	2	250	64,000	11,760	2
250	24,000	834	2	250	45,000	402	2
250	16,000	646	1	500	30,000	3,352	1
250	60,000	198	2	250	2,000	511	1
1,000	66,000	275	2	0	10,000	1,115	2
250	36,000	500	1	250	52,000	237	2
100	53,000	1,518	2	250	3,000	1,197	2
250	70,000	2,430	2	100	50,000	7,107	2
250	51,000	357	1	250	89,000	535	2
250	79,000	2,008	2	1,000	200,000	5,959	2
500	139,000	3,044	1	250	100,000	1,224	3
250	155,000	238	2	250	85,000	85,000*	1
250	150,000	3,244	2	250	103,000	2,358	2
250	98,000	850	2	250	110,000	31,243	2
250	100,000	198	2	500	110,000	1,488	1
100	110,000	110,000*	1	250	175,000	2,702	3
250	115,000	1,191	1	1,000	154,000	850	2
250	100,000	1,852	3	250	100,000	300	2
5,000	153,000	4,433	1	250	134,000	930	2
250	120,000	100	2	500	125,000	305	2
250	100,000	2,501	2	1,000	115,000	190	2
250	350,000	1,057	2	250	630,000	1,875	1
250	373,000	180	1	1,000	402,000	5,075	2
1,000	208,000	9,385	1	500	204,000	972	2
1,000	600,000	2,300	3	250	300,000	271	3
1,000	284,000	5,589	1	250	350,000	87	1
1,000	263,000	652	2	500	595,000	625	2
250	312,000	3,975	1	1,000	275,000	20,934	1
250	280,000	485	2	250	290,000	609	1
1,000	312,000	2,092	2	250	560,000	325	2
2,500	250,000	250,000*	1	1,000	371,000	6,012	1
250	300,000	250	2	1,000	362,000	860	2
500	625,000	1,305	3	250	317,000	2,720	2
1,000	319,000	6,729	3	500	6,817,000	1,040	3
500	9,214,000	185	2	1,000	3,010,000	48,762	1
1,000	3,000,000	22,930	3	5,000	6,023,000	20,576	1
1,000	800,000	498	3	250	700,000	230	2
500	838,000	990	2	1,000	1,000,000	200	2
250	1,400,000	5,491	3	500	1,442,000	1,247	1
1,000	1,500,000	1,185	3	1,000	2,000,000	10,000	2
500	36,819,000	6,032	2	1,000	2,526,000	4,525	3
250	1,282,000	13,775	2	500	65,065,000	16,981	2
250	1,000,000	150	3	1,000	1,236,000	4,911	2
1,000	6,127,000	4,536	2	1,000	5,000,000	81,692	2
100	1,140,000	298	3	250	2,275,000	21,447	2
1,000	1,910,000	335	2	1,000	2,700,000	992	2

*Building losses with asterisks next to them are losses capped by their respective insured building values (right censored.)

Appendix B: Exhibit 1

An S-Plus Program to Compute Maximum Likelihood Estimate of Parameters & Maximized Likelihood Statistic for Weibull Distribution

```
mydata<-TableA
m<-data.frame(mydata)
Weibull<-function(lamda, alfa, data = data.matrix)
        D <- data.matrix[,1]
       PL <- data.matrix[,2]
        y <- data.matrix[,3]
       z <- D+((y <PL)*y+(y >=PL)*PL)
    delta1<- (D==0)*(y <PL)
    delta2<- (D> 0)*(y <PL)
    delta3<- (D==0)*(y >=PL)
    delta4<- (D> 0)*(y >=PL)
    L1 <- alfa*lamda*(z^(alfa-1))*exp(-lamda*(z^alfa))
    L2 <-(alfa*lamda*(z^(alfa-1))*exp(-lamda*(z^alfa)))/exp(-lamda*(D^alfa))
    L3 <- exp( - lamda * (z^alfa))
    L4 <- exp( - lamda * (z^alfa))/exp( - lamda * (D^alfa))
    logL<- delta1*log(L1)+delta2*log(L2)+delta3*log(L3)+delta4*log(L4)
   -logL  }
min.Weibull<-ms(~Weibull(lamda,alfa), data=m, start
=list(lamda=1,alfa=.15))
min.Weibull
value: 899.802
parameters:
     lamda     alfa
 0.4484192 0.223073
formula:  ~  Weibull(lamda, alfa)
100 observations
call: ms(formula =  ~ Weibull(lamda, alfa), data = m, start = list(lamda
= 1, alfa = 0.15))
```

S-Plus is a statistical package produced by StatSci, a division of MathSoft, Inc., Seattle, Washington.

Weibull density is: $f(x;\lambda,\alpha) = \alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha})$

Appendix B: Exhibit 2

An S-Plus Program to Compute Maximum Likelihood Estimate of Parameters & Maximized Likelihood Statistic for a Family of Lognormal Distributions Based on "Mean" Model D

```
mydata<-TableA
m<-data.frame(mydata)
lognormal.model.D <- function(b0,b1,b2,b3,sigma, data=data.matrix)
  {  D <- data.matrix[,1]
    PL <- data.matrix[,2]
     y <- data.matrix[,3]
     z <- D+(y*(y<PL)+PL*(y>=PL))
  cnst <- data.matrix[,4]
    C1 <- cnst == 1
    C2 <- cnst == 2
     d <-D+(D == 0)*1
    mu <- b0+b1*log(PL)+b2*C1+b3*C2
      delta1 <- (D == 0)*(y < PL)
      delta2 <- (D > 0)*(y < PL)
      delta3 <- (D == 0)*(y >= PL)
      delta4 <- (D > 0)*(y >= PL)
      L1 <- dlnorm(z,mu,sigma)
      L2 <- dlnorm(z,mu,sigma)/(1-plnorm(d,mu,sigma))
      L3 <- 1-plnorm(z,mu,sigma)
      L4 <- (1-plnorm(z,mu,sigma))/(1-plnorm(d,mu,sigma))
      logL <-delta1*log(L1)+delta2*log(L2)+delta3*log(L3)+delta4*log(L4)
 -logL  }
min.model.D<-ms(~lognormal.model.D(b0,b1,b2,b3,sigma), data=m,
 start=list(b0=4.568, b1=0.238, b2=1.068, b3=0.0403, sigma=1.322))
min.model.D
value: 892.7099
parameters:
       b0        b1       b2         b3     sigma
 1.715296 0.3317345 2.154994 0.4105021 1.898501
formula:  ~  lognormal.model.D(b0, b1, b2, b3, sigma)
100 observations
call: ms(formula = ~ lognormal.model.D(b0, b1, b2, b3, sigma), data=m,
start =list(b0=4.568, b1=0.238, b2=1.068, b3=0.0403, sigma=1.322))
```

Appendix B: Exhibit 3

An S-Plus Program To Compute Maximum Likelihood Estimate of Parameters & Maximized Likelihood Statistic for a Family of Lognormal Distributions Based on "Scale" Model B

```
mydata<-TableA
m<- data.frame(mydata)
lognormal.Scale.model.B<- function(b0,b1,b2,mu, data=data.matrix)
   {  D <- data.matrix[,1]
      PL <- data.matrix[,2]
      y <- data.matrix[,3]
    cnst <- data.matrix[,4]
      z <- D + (y*(y < PL)+PL*(y >= PL))
      C1 <- cnst == 1
      C2 <- cnst == 2
      d <- D + (D == 0) * 1
      sigma <- b0+b1*C1+ b2* C2
      delta1 <- (D == 0)*(y < PL)
      delta2 <- (D > 0)*(y < PL)
      delta3 <- (D == 0)*(y >= PL)
      delta4 <- (D > 0)*(y >= PL)
      L1 <- dlnorm(z,mu,sigma)
      L2 <- dlnorm(z,mu,sigma)/(1 - plnorm(d,mu,sigma))
      L3 <- 1 - plnorm(z,mu,sigma)
      L4 <- (1 - plnorm(z,mu,sigma))/(1 - plnorm(d,mu,sigma))
      logL <-delta1*log(L1)+delta2*log(L2)+delta3*log(L3)+delta4*log(L4)
   -logL   }
 min.Scale.B<- ms(~lognormal.Scale.model.B(b0,b1,b2,mu), data=m,
+ start=list(b0=2,b1=0,b2=0,mu=6))
 min.Scale.B
value: 892.4242
parameters:
       b0       b1        b2       mu
 1.583642 1.324647 0.1066956 6.55098
formula:  ~  lognormal.Scale.model.B(b0, b1, b2, mu)
100 observations
call: ms(formula =  ~ lognormal.Scale.model.B(b0, b1, b2, mu), data = m,
start = list(b0 = 2, b1 = 0, b2 = 0, mu = 6))
```

Approximations of the Aggregate Loss Distribution

Dmitry E. Papush, Ph.D., FCAS,
Gary S. Patrik, FCAS, and Felix Podgaits

Approximations of the Aggregate Loss Distribution

Dmitry Papush, Ph.D., FCAS,
Gary Patrik, FCAS
Felix Podgaits

Abstract

Aggregate Loss Distributions are used extensively in actuarial practice, both in ratemaking and reserving. A number of approaches have been developed to calculate aggregate loss distributions, including the Heckman-Meyers method, Panjer method, Fast Fourier transform, and stochastic simulations. All these methods are based on the assumption that separate loss frequency and loss severity distributions are available.

Sometimes, however, it is not practical to obtain frequency and severity distributions separately, and only aggregate information is available for analysis. In this case the assumption about the shape of aggregate loss distribution becomes very important, especially in the "tail" of the distribution.

This paper will address the question of what type of probability distribution is the most appropriate to use to approximate an aggregate loss distribution.

Introduction

Aggregate loss distributions are used extensively in actuarial practice, both in ratemaking and reserving. A number of approaches have been developed to calculate aggregate loss distribution, including the Heckman-Meyers method, Panjer method, Fast Fourier transform, and stochastic simulations. All these methods are based on the assumption that separate loss frequency and loss severity distributions are available.

Sometimes, however, it is not practical to obtain frequency and severity distributions separately, and only aggregate information is available for analysis. In this case, the assumption about the shape of aggregate loss distribution becomes very important, especially in the "tail" of the distribution.

This paper will address the question what type of probability distribution is the most appropriate to use to approximate an aggregate loss distribution. We start with a brief summary of some important results that have been published about the approximations to the aggregate loss distribution.

Dropkin [3] and Bickerstaff [1] have shown that the Lognormal distribution closely approximates certain types of homogeneous loss data. Hewitt, in [6], [7], showed that two other positive distributions, the gamma and log-gamma, also provide a good fit.

Pentikainen [8] noticed that the Normal approximation gives acceptable accuracy only when the volume of risk business is fairly large and the distribution of the amounts of the individual claims is not too heterogeneous. To improve the results of Normal approximation, the NP-method was suggested. Pentikainen also compared the NP-method with the Gamma approximation. He concluded that both methods give good accuracy when the skewness of the aggregate losses is less than 1, and neither Gamma nor NP is safe when the skewness of the aggregate losses is greater than 1.

Seal [9] has compared the NP method with the Gamma approximation. He concluded that the Gamma provides a generally better approximation than NP method. He also noted that the superiority of the Gamma approximation is even more transparent in the "tail" of the distribution.

Sundt [11] in 1977 published a paper on the asymptotic behavior of the compound claim distribution. He showed that under some special conditions, if the distribution of the number of claims is Negative Binomial, then the distribution of the aggregate claims behaves asymptotically as a gamma-type distribution in its tail. A similar result is described in [2] (Lundberg Theorem, 1940). The theorem states that under certain conditions, a negative binomial frequency leads to an aggregate distribution, which is approximately Gamma.

The skewness of the Gamma distribution is always twice its coefficient of variation. Since the aggregate loss distribution is usually positively skewed, but does not always have skewness double its coefficient of variation, adding a third parameter to the Gamma was suggested by Seal [9]. However, this procedure may give positive probability to negative losses. Gendron and Crepeau [4] found that, if severity is Inverse Gaussian and frequency is Poisson, the Gamma approximation produce reasonably accurate results and is superior to the Normal, N-P and Escher approximations when the skewness is large.

In 1983, Venter [12] suggested the Transformed Gamma and Transformed Beta distributions to approximate the aggregate loss distributions. These gamma-type distributions, allowing some deviation from the Gamma, are thus appealing candidates.

This paper continues the research into the accuracy of different approximations of the aggregate loss distribution. However, there are two aspects that differentiate it from previous investigations.

First, we have restricted our consideration to two-parameter probability distributions. While adding the third parameter generally improves accuracy of approximation, observed samples are usually not large enough to warrant a reliable estimate of an extra, third, parameter.

Second, all prior research was based upon theoretical considerations, and did not consider directly the goodness of fit of various approximations. We are using a different approach, building a large simulated sample of aggregate losses, and then directly testing the goodness of fit of various approximations to this simulated sample.

Description of the Method Used

The ideal method to test the fit of a theoretical distribution to a distribution of aggregate losses would be to compare the theoretical distribution with an actual, statistically representative, sample of observed values of the aggregate loss distribution. Unfortunately, there is no such sample available: no one insurance company operates in an unchanged economic environment long enough to observe a representative sample of aggregate (annual) losses. Economic trend, demography, judicial environment, even global warming, all impact the insurance marketplace and cause the changes in insurance losses. Considering periods shorter than a year does not work either because of seasonal variations.

Even though there is no historical sample of aggregate losses available, it is possible to create samples of values that could be aggregate insurance losses under reasonable frequency and severity assumptions. Frequency and severity of insurance losses for major lines of business are being constantly analyzed by individual insurance companies and rating agencies. The results of these analyses are easily available, and of a good quality. Using these data we can simulate as many aggregate insurance losses as necessary and then use these simulated losses as if they were actually observed: fit a probability distribution to the sample and test the goodness of fit. The idea of this method is similar to the one described by Stanard [10]: to simulate results using reasonable underlying distributions, and then use the simulated sample for analysis.

Our analysis involved the following formal steps:

1. Choose severity and number of claims distributions;
2. Simulate the number of claims and individual claim amounts, and calculate the corresponding aggregate loss;
3. Repeat many times (5,000) to obtain a sample of aggregate losses;
4. For different probability distributions, estimate their parameters, using the simulated sample of aggregate losses;
5. Test the goodness of fit for the various probability distributions.

Selection of Frequency and Severity Distributions

Conducting our study, we kept in mind that the aggregate loss distribution could potentially behave very differently, depending on the book of business covered. Primary insurers usually face massive frequency (large number of claims), with limited fluctuation in severity (buying per occurrence excess reinsurance). To the contrary, an excess reinsurer often deals with low frequency, but a very volatile severity of losses. To reflect possible differences, we tested several scenarios that are summarized in the following table.

Scenario #	Type of Book of Business	Expected Number of Claims	Per Occurrence Limit	Type of Severity Distribution
1	Small Primary, Low Retention	50	\$0 – 250K	5 Parameter Pareto
2	Large Primary, Low Retention	500	\$0 – 250K	5 Parameter Pareto
3	Small Primary, High Retention	50	\$0 – 1000K	5 Parameter Pareto
4	Large Primary, High Retention	500	\$0 – 1000K	5 Parameter Pareto
5	Working Excess	20	\$750K xs \$250K	5 Parameter Pareto
6	High Excess	10	\$4M xs \$1M	5 Parameter Pareto
7	High Excess	10	\$4M xs \$1M	Lognormal

Number of claims distribution for all scenarios was assumed to be Negative Binomial. Also, we used Pareto for the severity distribution in both primary and working excess layers. In these (relatively) narrow layers, the shape of the severity distribution selected has a very limited influence on the shape of the aggregate distribution. In a high excess layer, where the type of severity distribution can make a material difference, we tested two severity distributions: Pareto and Lognormal. More details on parameter selection for the frequency and severity distribution can be found in the exhibits that summarize our findings for each scenario.

Distributions Used for the Approximation of Aggregate Losses

As we discussed before, we concentrated our study on two-parameter distributions. Basically, we tested three widely used two-parameter distributions, to test their fits to the aggregate loss distributions constructed in each of the seven scenarios. Each of these three distributions was an appealing candidate to provide a good approximation. The following table lists the three distributions used.

Type of Distribution	Parameters	Probability Density Function	Mean	Variance
Normal	μ $\sigma > 0$	$f(x) = 1/(\sigma\sqrt{2\pi})$ * $\exp(-(x - \mu)^2/(2\sigma^2))$	μ	σ^2
Lognormal	μ $\sigma > 0$	$f(x) = 1/(\sigma x\sqrt{2\pi})$ * $\exp(-(\ln x - \mu)^2/(2\sigma^2))$	$\exp(\mu + \sigma^2/2)$	$\exp(2\mu + \sigma^2)$ * $[\exp(\sigma^2) - 1]$
Gamma	$\alpha > 0$ $\beta > 0$	$f(x) = 1/(\Gamma(x))$ * $\beta^{-\alpha}x^{\alpha-1}\exp(-x/\beta)$	$\alpha\beta$	$\alpha\beta^2$

A Normal distribution appears to be a reasonable choice, at least when the expected number of claims is sufficiently large. One would expect a Normal approximation to work in this case because of the Central Limit Theorem (or, more precisely, its generalization for random sums; see, for instance, [5]). As we shall see, however, to make this happen, the expected number of claims must be extremely large.

A Lognormal distribution has been used extensively in actuarial practice to approximate both individual loss severity and aggregate loss distributions ([1], [3]). A Gamma distribution also has been claimed by some authors ([6], [9]) to provide a good fit to aggregate losses.

Parameter Estimates and Tests of Goodness of Fit

Initially we used both the Maximum Likelihood Method and the Method of Moments to estimate parameters for the approximating distributions. The parameter estimates obtained by the two methods were reasonably close to each other. Also, the distribution based on the parameters obtained by the Method of Moments provided a better fit than the one based on the parameters obtained by the Maximum Likelihood Method. For these reasons we have decided to use the Method of Moments for parameter estimates.

Once the simulated sample of aggregate losses and the approximating distributions were constructed, we tested the goodness of fit. While the usual "deviation" tests (Kolmogorov – Smirnov and χ^2-test) provide a general measurement of how close two distributions are, they can not help to determine if the distributions in question systematically differ from each other for a broad range of values, especially in the "tail". To pick up such differences, we used two tests that compare two distributions on their full range.

The Percentile Matching Test compares the values of distribution functions for two distributions at various values of the argument up to the point when the distribution functions effectively vanish. This test is the most transparent indication of where two distributions are different and by how much.

The Excess Expected Loss Cost Test compares the conditional means of two distributions in excess of different points. It tests values $E[X - x \mid X > x] * \text{Prob}\{X > x\}$. These values represent the loss cost of the layer in excess of x if X is the aggregate loss variable. The excess loss cost is the most important variable for both the ceding company and reinsurance carrier, when considering stop loss coverage, aggregate deductible coverage, and other types of aggregate reinsurance transactions.

Results and Conclusions

The four exhibits at the end of the paper document the results of our study for each of the seven scenarios described above. The exhibits show the characteristics of the frequency and severity distributions selected for each scenario, estimators for the parameters of the three approximating distributions, and the results of the two goodness-of-fit tests.

The results of the study are quite uniform: for all seven scenarios the Gamma distribution provides a much better fit than the Normal and Lognormal. In fact, both Normal and Lognormal distributions show unacceptably poor fits, but in different directions.

The Normal distribution has zero skewness and, therefore, is too light in the tail. It could probably provide a good approximation for a book of business with an extremely large expected number of claims. We have not considered such a scenario however.

In contrast, the Lognormal distribution is overskewed to the right and puts too much weight in the tail. The Lognormal approximation significantly misallocates the expected losses between excess layers. For the Lognormal approximation, the estimated loss cost for a high excess layer could be as much as 1500% of its true value.

On the other hand, the Gamma approximation performs quite well for all seven scenarios. It still is a little conservative in the tail, but not as conservative as the Lognormal. This level of conservatism varies with the skewness of the underlying severity distribution, and reaches its highest level for scenario 2 (Large Book of Business with Low Retention). When dealing with this type of aggregate distribution, one might try other alternatives.

As the general conclusion of this study, we can state that the Gamma distribution gives the best fit to aggregate losses out of the three considered alternatives for the cases considered. It can be recommended to use the Gamma as a reasonable approximation when there is no separate frequency and severity information available.

Bibliography.

1. Bickerstaff, D. R. *Automobile Collision Deductibles and Repair Cost Groups: The Lognormal Model*, PCAS LIX (1972), p. 68.

2. Cramer, H. *Collective Risk Theory*, The Jubilee Volume of Forsakringsaktiebolaget Skandia, 1955.

3. Dropkin, L. B. *Size of Loss Distributions in Workmen's Compensation Insurance*, PCAS LI (1964), p. 68.

4. Gendron, M., Crepeau H. *On the computation of the aggregate claim distribution when individual claims are Inverse Gaussian*, Insurance: Mathematics and Economics, 8:3, 1989, p. 251.

5. Gnedenko B.V., Korolev V.Yu. *Random Summation: Limit Theorems and Applications*. CRC Press, 1996.

6. Hewitt, C.C. *Distribution by Size of Risk – A Model*, PCAS LIII (1966), p. 106.

7. Hewitt, C.C. *Loss Ratio Distributions – A Model*, PCAS LIV (1967), p. 70.

8. Pentikainen, T. *On the Approximation of the total amount of claims*, ASTIN Bulletin, 9:3, 1977, p. 281.

9. Seal, H. *Approximations to risk theory's F(X, t) by means of the gamma distribution*, ASTIN Bulletin, 1977.

10. Stanard, J.N., *A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques*, PCAS LXXII (1985), p. 124.

11. Sundt, B. *Asymptotic behavior of Compound Distributions and Stop-Loss Premiums*, ASTIN Bulletin, 13:2, 1982, p. 89.

12. Venter, G. *Transformed Beta and Gamma Distributions and aggregate losses*, PCAS LXX (1983).

Exhibit 1

Scenario 1.

Frequency:	Negative Binomial	
Expected Number of Claims		50
Severity:	5 Parameter Truncated Pareto	
Expected Severity		13,511
Per Occurrence Limit		250,000

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	13.347	Mu	691,563	Alpha	4.521
Sigma	0.447	Sigma	325,246	Beta	152,965
Mean	691,563			Mean	691,563

Percentile matching

	P(X>x)			
x	**Empirical**	Lognormal	Normal	**Gamma**
500,000	**69.36%**	69.22%	72.21%	**68.90%**
750,000	**38.06%**	34.27%	42.87%	**37.02%**
1,000,000	**16.48%**	14.72%	17.15%	**16.16%**
1,250,000	**6.16%**	6.08%	4.30%	**6.11%**
1,500,000	**1.94%**	2.53%	0.65%	**2.09%**
1,750,000	**0.62%**	1.07%	0.06%	**0.66%**
2,000,000	**0.06%**	0.47%	0.00%	**0.20%**

Expected Loss costs

E[X-x| X>x] * P(X>x)			
Empirical	Lognormal	Normal	**Gamma**
237,751	227,011	178,648	**234,823**
104,504	100,316	43,996	**103,823**
38,636	42,118	5,123	**40,019**
12,293	17,660	245	**13,924**
3,518	7,553	4	**4,483**
870	3,323	0	**1,359**
111	1,507	0	**393**

Scenario 2.

Frequency:	Negative Binomial	
Expected Number of Claims		500
Severity:	5 Parameter Truncated Pareto	
Expected Severity		13,511
Per Occurrence Limit		250,000

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	15.740	Mu	6,922,204	Alpha	47.462
Sigma	0.144	Sigma	1,004,786	Beta	145,849
Mean	6,922,204			Mean	6,922,204

Percentile matching

	P(X>x)			
x	**Empirical**	Lognormal	Normal	**Gamma**
6,000,000	**82.48%**	82.07%	82.06%	**81.94%**
7,000,000	**44.92%**	44.05%	46.91%	**45.01%**
8,000,000	**13.74%**	14.13%	14.17%	**14.25%**
9,000,000	**2.64%**	2.94%	1.93%	**2.63%**
9,500,000	**1.02%**	1.18%	0.52%	**0.94%**
10,000,000	**0.28%**	0.44%	0.11%	**0.30%**
10,500,000	**0.02%**	0.16%	0.02%	**0.09%**

Expected Loss costs

E[X-x| X>x] * P(X>x)			
Empirical	Lognormal	Normal	**Gamma**
1,009,130	1,001,072	836,942	**1,007,562**
362,107	362,956	170,371	**363,947**
83,509	89,015	10,310	**83,937**
11,978	15,524	137	**12,315**
3,521	5,838	8	**4,024**
586	2,072	0	**1,192**
16	699	0	**322**

Exhibit 2

Scenario 3.

Frequency:	Negative Binomial	
Expected Number of Claims		50
Severity:	5 Parameter Truncated Pareto	
Expected Severity		18,991
Per Occurrence Limit		1,000,000

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	13.590	Mu	958,349	Alpha	2.265
Sigma	0.605	Sigma	636,775	Beta	423,106
Mean	958,349			Mean	958,349

Percentile matching

		P(X>x)		
x	Empirical	Lognormal	Normal	Gamma
1,000,000	**38.68%**	35.47%	47.39%	**38.70%**
1,500,000	**18.28%**	14.84%	19.75%	**17.27%**
2,000,000	**6.92%**	6.44%	5.09%	**7.08%**
2,500,000	**2.82%**	2.95%	0.77%	**2.76%**
2,750,000	**1.54%**	2.04%	0.24%	**1.69%**
3,000,000	**0.92%**	1.43%	0.07%	**1.03%**
3,250,000	**0.42%**	1.01%	0.02%	**0.62%**
3,500,000	**0.28%**	0.73%	0.00%	**0.37%**

Expected Loss costs

	E[X-x| X>x] * P(X>x)		
Empirical	Lognormal	Normal	Gamma
233,797	212,405	110,782	**228,287**
94,548	94,109	13,815	**94,254**
35,445	44,012	692	**36,798**
12,438	21,761	13	**13,826**
7,085	15,599	1	**8,382**
4,021	11,313	0	**5,052**
2,534	8,296	0	**3,029**
1,697	6,145	0	**1,807**

Scenario 4.

Frequency:	Negative Binomial	
Expected Number of Claims		500
Severity:	5 Parameter Truncated Pareto	
Expected Severity		18,991
Per Occurrence Limit		1,000,000

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	16.065	Mu	9,685,425	Alpha	23.564
Sigma	0.204	Sigma	1,995,223	Beta	411,021
Mean	9,685,425			Mean	9,685,425

Percentile matching

		P(X>x)		
x	Empirical	Lognormal	Normal	Gamma
10,000,000	**40.96%**	39.79%	43.74%	**41.12%**
12,000,000	**12.50%**	12.44%	12.30%	**12.59%**
14,000,000	**2.18%**	2.81%	1.53%	**2.43%**
15,000,000	**0.88%**	1.23%	0.39%	**0.92%**
16,000,000	**0.36%**	0.52%	0.08%	**0.32%**
17,000,000	**0.12%**	0.21%	0.01%	**0.10%**
18,000,000	**0.06%**	0.08%	0.00%	**0.03%**

Expected Loss costs

	E[X-x| X>x] * P(X>x)		
Empirical	Lognormal	Normal	Gamma
650,476	651,609	283,657	**654,235**
160,831	165,420	14,936	**151,977**
22,879	33,145	166	**24,231**
8,930	13,941	9	**8,544**
3,160	5,689	0	**2,799**
1,060	2,268	0	**857**
186	888	0	**247**

Scenario 5.

Frequency:	Negative Binomial	
Expected Number of Claims		20
Severity:	5 Parameter Truncated Pareto	
Expected Severity		315,640
Per Occurrence Excess Layer		$750K x $250K
Skewness		0.416

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	15.571	Mu	6,306,951	Alpha	5.301
Sigma	0.416	Sigma	2,739,428	Beta	1,189,872
Mean	6,306,951			Mean	6,306,951

Percentile matching

	P(X>x)			
x	Empirical	Lognormal	Normal	Gamma
6,000,000	**50.26%**	46.50%	54.46%	**48.72%**
8,000,000	**24.48%**	21.77%	26.83%	**23.81%**
10,000,000	**9.70%**	9.40%	8.88%	**9.87%**
12,000,000	**3.28%**	3.96%	1.88%	**3.63%**
14,000,000	**1.00%**	1.68%	0.25%	**1.22%**
16,000,000	**0.26%**	0.72%	0.02%	**0.38%**
20,000,000	**0.04%**	0.14%	0.00%	**0.03%**

Expected Loss costs

E[X-x| X>x] * P(X>x)			
Empirical	Lognormal	Normal	Gamma
1,225,433	1,173,911	682,504	**1,218,440**
503,151	515,230	120,368	**511,426**
174,623	219,525	9,991	**191,144**
54,155	93,588	355	**65,274**
14,274	40,508	5	**20,761**
3,491	17,921	0	**6,238**
772	3,779	0	**494**

Scenario 6.

Frequency:	Negative Binomial	
Expected Number of Claims		10
Severity:	5 Parameter Truncated Pareto	
Expected Severity		1,318,316
Per Occurrence Excess Layer		$4M x $1M
Skewness		1.887

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	16.006	Mu	12,985,319	Alpha	0.901
Sigma	0.864	Sigma	13,683,648	Beta	14,419,533
Mean	12,985,319			Mean	12,985,319

Percentile matching

		P(X>x)		
x	**Empirical**	Lognormal	Normal	**Gamma**
15,000,000	**31.80%**	27.46%	44.15%	**31.13%**
20,000,000	**22.42%**	17.57%	30.41%	**21.61%**
25,000,000	**15.32%**	11.70%	19.00%	**15.05%**
30,000,000	**10.56%**	8.06%	10.69%	**10.50%**
40,000,000	**5.36%**	4.15%	2.42%	**5.14%**
50,000,000	**2.52%**	2.32%	0.34%	**2.52%**
60,000,000	**1.10%**	1.38%	0.03%	**1.24%**

Expected Loss costs

E[X-x| X>x] * P(X>x)			
Empirical	Lognormal	Normal	**Gamma**
4,359,267	3,731,938	1,991,361	**4,315,503**
3,026,841	2,628,509	806,980	**3,011,961**
2,094,021	1,908,864	271,738	**2,105,672**
1,455,987	1,421,737	74,986	**1,473,927**
689,160	837,602	3,009	**724,182**
324,128	525,046	49	**356,764**
152,034	345,040	0	**176,096**

Scenario 7.

Frequency:	Negative Binomial	
Expected Number of Claims		10
Severity:	Lognormal	
Expected Severity		2,166,003
Per Occurrence Excess Layer		$4M x $1M
Skewness		1.190

Method of Moments estimated parameters for:

Lognormal		Normal		Gamma	
Mu	16.601	Mu	20,233,595	Alpha	1.786
Sigma	0.667	Sigma	15,141,348	Beta	11,330,681
Mean	20,233,595			Mean	20,233,595

Percentile matching

		P(X>x)		
x	**Empirical**	Lognormal	Normal	**Gamma**
20,000,000	**42.16%**	37.60%	50.62%	**40.65%**
25,000,000	**30.66%**	25.76%	37.65%	**29.44%**
30,000,000	**21.52%**	17.77%	25.95%	**20.99%**
40,000,000	**10.64%**	8.76%	9.59%	**10.31%**
50,000,000	**5.24%**	4.55%	2.47%	**4.91%**
60,000,000	**2.06%**	2.48%	0.43%	**2.29%**
70,000,000	**0.76%**	1.41%	0.05%	**1.05%**

Expected Loss costs

E[X-x| X>x] * P(X>x)			
Empirical	Lognormal	Normal	**Gamma**
5,984,377	5,371,757	3,116,920	**5,861,977**
4,167,564	3,806,611	1,488,582	**4,122,201**
2,877,740	2,731,389	615,453	**2,872,036**
1,299,737	1,462,583	65,324	**1,364,938**
546,755	822,414	3,471	**635,179**
204,979	482,606	88	**291,051**
69,923	293,789	1	**131,796**

Extended Warranty Ratemaking

L. Nicholas Weltmann, Jr., FCAS, MAAA and
David Muhonen

EXTENDED WARRANTY RATEMAKING

TABLE OF CONTENTS

Abstract

The warranty business is a relatively new line of insurance in the property-casualty market. For the most part insurance coverage for warranties, extended warranties and service contract reimbursement policies has been introduced over the last thirty years. There is great opportunity in this line of business for the pricing actuary. It is an area where one can use his imagination and creativity in developing actuarially sound models to price and evaluate warranty business.

This paper starts with auto extended warranty ratemaking, where there is usually plenty of data to use the traditional actuarial approaches to ratemaking. From there the paper discusses a non-traditional rate-making approach when historical experience is not available. This "back-to-basics" approach focuses on developing the pure premium by independently deriving frequency and severity. The next topic is the inclusion of unallocated loss adjustment expense (ULAE) into the pricing equation. In this line of business, because of the long-term commitments, ULAE must be carefully analyzed and provided for. Lastly, the paper discusses a number of pricing pitfalls to avoid. Some of these errors have been made by the author, and it is in the hopes of exposing these pitfalls that they can be avoided by others.

Introduction

The warranty business is a relatively new line of business to the property/casualty market. It is generally within the last thirty years that insurance coverage has become an integral method to transfer this risk. Warranty coverage is basically mechanical breakdown insurance; if a product does not work due to some mechanical or component failure and it is covered under a warranty contract, than the product is either repaired or replaced, depending on the type of coverage in force.

Relatively speaking, there is very little actuarial literature on the topic of warranty business in general. Several that come to mind are the 1994 Proceedings paper by Roger Hayne, "Extended Service Contracts" and two papers in the 1993 CAS Forum Ratemaking Call Papers, "A Pricing Model for New Vehicle Extended Warranties" by Joseph S. Cheng and Stephen J. Bruce, and "The Use of Simulation Techniques in Addressing Auto Warranty Pricing and Reserving Issues" by Simon J. Noonan. Some of the topics addressed in those papers will be touched on in this paper.

The pricing of a warranty product lends itself to the pricing actuary's expertise. It is generally a line that has predictable frequencies and severities, given a credible amount of data. On the auto warranty class, there is usually a great deal of data available to analyze using traditional actuarial methods. Other product areas do not have large amounts of data and the actuary is forced to develop a price by deriving a value for frequency and severity.

The warranty market today can be divided up into five basic segments, each with its own set of distinguishing characteristics. These segments would be the automobile service contracts, commercial warranties (example; policies covering business equipment), home warranties (example; public service policies covering furnaces and air conditioners), retail warranties (example; policies covering VCRs) and Original Equipment Manufacturers (OEM) warranties. In this paper we will discuss auto extended warranty ratemaking and OEM warranty ratemaking, as well as several general topics which touch all areas.

VEHICLE EXTENDED WARRANTY

The auto extended warranty concept dates back to the early 1970's. Prior to that the only warranties on automobiles were the manufacturer's warranties on new vehicles, which were generally limited to 12 months or 12,000 miles. Used cars were usually sold with no warranty.

In the early 1970's a few independent companies, generally not insurance companies but third party administrators (TPA) began to offer limited warranties on used cars. Soon there were a number of companies offering one, two and three year terms for these warranties.

Eventually these independents recognized another market could be extending the warranty beyond what was offered by the manufacturer. Covering new vehicles appeared to be a great cash flow bonanza, as the money for the coverage was paid up front, while claims would be delayed by the year's coverage under the manufacturer's warranty. Interest rates were very high in the early and middle of the 1980's, and investors were lured by the promise of high returns. Manufacturers began to offer their own extended warranties, forcing independent TPAs out or to reduce pricing. Some of these TPAs were backed by insurance companies; many were not.

The late 1980's saw a turmoil in this business as pricing on new vehicle service contracts (VSC) was woefully inadequate. During this time the manufacturers also began to lengthen the term of the underlying warranty to three years or thirty-six thousand miles. This posed an immediate pricing problem. Purchasers of an extended warranty would expect the pricing to go down as the manufacturer now covered more claims. However, actuarial studies indicated double digit rate increases necessary. Interest rates also were coming down, lowering the investment income.

TPAs that raised rates lost much of their volume almost overnight, as dealers had a choice of the manufacturers' or other independents' products. However, a number of independents did survive this period. Most of these are either owned by or closely affiliated with an insurance company for security reasons, as long-term promises of vehicle service are being made. The manufacturers control about 70% of the new vehicle extended warranty market with the independents sharing the rest. The independents have a greater share of the used vehicle market.

Insurance companies play an important role in the selling of the extended warranty product. The extended warranty is an after-market product, that is, the dealer and consumer will generally conclude the purchase of the vehicle before introducing the availability of the extended warranty. If the dealer is successful in selling the consumer an extended warranty or service contract, the dealer has then committed to a long-term relationship to service that vehicle.

In most states, the extended warranty service contract is not considered insurance and is not regulated by the insurance department. It is simply a contract between the dealer and the car buyer and is covered under contract law. What is considered insurance by most states and is regulated by the various insurance departments is the Service Contract Reimbursement Policy (SCRIP). If the dealer chooses to sell an independent TPA's VSC, the dealer needs to assure himself that the TPA will be there to fulfill the promises made to the consumer. The consumer also must satisfy himself that should he move from the area or the dealer goes out of business, covered repairs will still be made. The TPA must therefore show that he is secure; most TPA's, through an insurance company, therefore provide a SCRIP to the dealer. This SCRIP provides a guarantee to the dealer and the consumer that if a covered repair is necessary it will be done, either at the selling dealer or at an authorized repair shop.

The vehicle service contract

The vehicle service contract (VSC) has a number of options in terms of limits and coverage. The predominate products will be discussed here. The discussion will be broken into three segments; used vehicles, new vehicles and near-new vehicles. Used vehicles are those which are being resold to the consumer by a dealer and which no longer are covered by the manufacturer's warranty. New vehicles are those which have had no previous owners and have the full protection of the manufacturer's warranty. Near-new vehicles are those that have had a previous owner and are being resold by the dealer with some protection still under the manufacturer's warranty.

Used Vehicle Service Contracts - Limits

a. One-year term – The VSC coverage is limited to one year from purchase of vehicle. Mileage on the vehicle at time of purchase is also used as an eligibility factor, i.e., a vehicle with mileage beyond a certain limit will not be eligible for an extended warranty.

b. Two-year term – This VSC coverage is limited to two years from the purchase of the vehicle. Again a mileage limit as described above is in place, but it is usually lower than the one-year eligibility as the coverage lasts longer.

c. Three-year term – This VSC coverage is limited to three years from the time of purchase with an eligibility mileage limit in place. Again, this eligibility limit would normally be lower than that for the two-year term.

New Vehicle Service Contracts – Limits

The limits on a new VSC are almost always a combination of years and mileage. The most popular combinations are usually in multiples of whole years (5,6 or 7) and multiples of 10,000 miles, from 60,000 to 100,000. An example of how this is shown would be 5/100,000, which represents 5 years or 100,000 miles, whichever comes first. At one time an option for unlimited mileage was offered, but industrywide experience was so poor that this option is now very seldom seen. Coverage starts upon the purchase of the vehicle.

Near-new Vehicle Service Contracts – Limits

These limits would normally be expressed as those shown for new VSCs. In fact, until recently this group was not separated from the "new" grouping. A new VSC would be sold to a consumer as long as there was still coverage under the manufacturer's warranty, the theory being that there was very little exposure to loss anytime during the period under which the vehicle was covered by the manufacturer. Upon analysis, however, it was found that loss costs were higher for new VSCs sold 18 months after coverage started under the manufacturer's warranty than for new VSCs sold on vehicles within that 18 month period.

We initially began to study the loss costs of this group because we noted that a program which we underwrote for motorcycles had much higher loss experience for older bikes which were grandfathered into the program. These older motorcycles were only eligible for the new program if they had been purchased no more than one year prior to the inception of the program. The resulting loss costs on these bikes were significantly higher than the rest of the program; we guessed that there was some type of adverse selection taking place.

If adverse selection was taking place in our motorcycle program where we provided an option to purchase an VSC more than a year after the bike was bought, then it would be reasonable to assume that the same adverse selection was taking place when a car owner purchased a VSC more than a year after he bought the car. As noted above our subsequent analysis of the near-new group showed significantly higher loss costs in comparison to the new group, and we therefore created the near-new group with higher rates.

Before the two-year lease option became popular, this group of vehicles was very small. However, this group has grown substantially over the last five years as the two year lease became predominant. Remember, the most prominent manufacturer's warranty is now 3/36,000, so a vehicle coming off a two-year lease still has up to a year of underlying coverage, depending on mileage.

Coverage offered under a vehicle service contract

Coverage under the VSC is for mechanical breakdown due to failure of a covered component only, and perhaps some incidental coverage such as rental reimbursement and towing when a covered mechanical breakdown has occurred. No physical damage due to other perils is covered. For instance, an engine breakdown caused when a vehicle is caught in a flood is not covered.

There are usually several options available in terms of coverage. There are a myriad of components that make up the automobile, with some obviously being more essential to the actual running of the auto than others. Basic coverage would normally cover the powertrain of the vehicle, such as the engine and transmission. Other options could be offered, up to "bumper-to-bumper" which pretty much covers everything in and on the car.

Vehicle Service Contract Ratemaking

Before discussing the actual ratemaking for VSCs, it is important to understand the makeup of the total price paid by the ultimate consumer, the purchaser of the vehicle. The total price is comprised of:

$$P = I + A + T + M;$$

where P = total price,

I = Insurer cost,

A = Agents commissions,

T = TPA administrative costs, and

M = Dealer markup.

To clarify, let us build the ultimate price to the consumer from the bottom up. First, the insurer determines the expected loss costs and adds any internal company expenses. This is passed to the TPA as the insurer cost. The TPA has administrative costs (underwriting, claims, systems, etc) which then get added on to the insurer cost. For the most part the TPA has an independent agency force in place to sell the SCRIP to the dealer, thus agent's commissions must also be included. (Note that as we pointed out earlier, the dealer sells the consumer a VSC, which is not typically considered insurance, and thus the dealer is not an insurance agent.) All of the above costs make up what is called the dealer's cost, to which the dealer then adds whatever markup he can to arrive at the total price. Since this markup is not regulated in any state but Florida, total price for the same VSC can vary from consumer to consumer, depending on the negotiating skills of the buyer and seller.

Dealer markup is not regulated in any state but Florida, and therefore is not included as a cost in filed rates anywhere but Florida. The remaining costs, however, may or may not be included in filed rates. Some companies file rates which only include insurer costs (I); the TPA will then collect a fee per VSC (T + A) from the dealer, which he will then have to use to pay the TPA's expenses as well as any commissions to his distribution force, if any. The filed rates may include I +T + A, in which case the insurer will pay out a commission to the TPA equal to T+A. In Florida, the filed rates include all costs. While these different scenarios do not present a problem for ratemaking, it does cause difficulty if one is trying to do a competitive rating study among various companies, as unless the costs included in the ratemaking are known, comparisons are almost worthless.

Insurer costs (I) are the next item of evaluation. Insurer costs are made up of expected loss costs and the insurer's expenses. The expected loss costs are a function of many variables, including but not limited to:

a. Manufacturer (Asian, US, European)

b. Coverage option

c. Make (Ford, Toyota, etc) and model (Explorer, Corolla, etc)

d. Term limit option

e. Mileage limit option

f. Deductible option

g. Underlying warranty (manufacturer's warranty)

h. Special factors (four-wheel drive, commercial use, advanced technology for example).

The company must decide what loss cost variables they would like to include in the ratemaking; the above would be a pretty standard method to analyze data. As the variables above are all important elements that differentiate rates, it is important that the data be captured in the same detail. It is also important that the data be analyzed on a policy year basis. Because of the multi-year terms of the policies, it is important to match the losses to the policies that generated those losses. It also avoids any distortions caused by improper earning of the premium.

The earning of the premium for a warranty product is not straightforward. In general, premium is earned over the policy period to reflect the exposure to loss during that policy period. For an annual policy the premium is usually earned pro rata as losses are assumed to be uniformly distributed over the policy period. This is not true in the extended warranty coverage.

For used VSCs losses generally come in faster than a pro rata distribution. A useful rule of thumb is that half of the losses have emerged when the term is one-third expired, and two-thirds of the losses are emerged when the term is half done. For example, on a two-year term used VSC, two-thirds of the losses have emerged one year into the term. One primary reason for this accelerated loss pattern is that mechanical problems on used vehicles can occur pretty quickly after the sale. Sometimes a used car dealer will use the extended warranty as a maintenance program. (This will be discussed later in the dealer management section.) For used VSCs, the premiums should be earned accordingly.

On new VSCs the earning is somewhat trickier. First, very few losses are expected under the extended warranty while the underlying warranty has not expired. The only losses during this period would be towing or rental expenses over and beyond what the underlying covers. Once the underlying warranty has expired, losses emerge on the extended warranty cover. As the frequency and severity of repairs are expected to increase during the remainder of the service contract we would envision an ever-increasing loss payout pattern. This type of pattern is well described by the reverse sum of the digits function (see Exhibit E for definition and formula), and this pattern is often used.

However, in actuality, while loss emergence does accelerate for a period of time after the expiration of the underlying warranty, this emergence slows down considerably towards the end of the term. This variable is sometimes called the attrition factor. Several things may happen during the life of the VSC; the mileage limit could be hit before the term limit, the car may be sold and the warranty not transferred, the owner voids the warranty by poor maintenance, or even the owner just doesn't keep track of the warranty contract. In any event, this attrition factor does exist, and it causes the loss payout pattern to take an "S" shape, slow starting out, grows quickly in the middle and slows down at the end. Premiums should be earned in the same fashion.

The loss payout patterns are direct byproducts of the actuarial analysis of the policy year loss triangles. The actuary decides at what level the earnings should be done, and has the data collected in these levels. For instance, earnings may be done by term and mileage, so premiums and losses would be segregated into term and mileage subsets by policy year.

Losses are developed to ultimate using a variety of methods. Because the loss emergence is low in the beginning of the contract period, more recent policy years benefit from the use of the Bornhuetter/Ferguson (B/F)* and the Stanard/Buhlmann (S/B)** methods in addition to simply multiplying the selected loss development factor by the emerged losses. It is also valuable to use average claim costs to develop ultimate losses (See Exhibit A). Note that for more recent years the paid loss projection is erratic as there are few emerged losses.

We also calculate a pure premium projection of ultimate losses (columns 13-15 in Exhibit A.) We use the B/F annual projection to get an ultimate pure premium per contract (column 13.) The B/F projection is used as its values are between the paid and the S/B projections, and thus we hope to be neither too optimistic nor too conservative. In column 14 we convert the annual pure premium into a running cumulative pure premium. In this way we incorporate mature years' pure premiums which have minimal actuarial adjustments along with the more recent years' pure premiums which are very dependent on actuarial assumptions on development. We then multiply the number of contracts written (column 2) by the cumulative pure premium to obtain the pure premium projection in column 15.

* For definition and explanation of the B/F method, please see Foundations of Casualty Actuarial Science, pages 210-214.

** For definition and explanation of the S/B method, please see Foundations of Casualty Actuarial Science, pages 352-354.

Of course, the other actuarial adjustments must also be made. Premiums must also be developed to ultimate as well as put on current rate level, and losses must be trended from the midpoint of the experience period to the midpoint of the proposed policy period. Individual policy years are then averaged and compared against the expected loss ratio to compute the required rate level indication.

LOSS TREND

Loss trend is a function of change in frequency vs a change in severity. For auto warranty business, normally the frequency is high and the severity is low. Frequency is affected by changes to the underlying manufacturers' warranties, the quality of the vehicles, the changing mix of business, and the dealers' service departments' propensity to use the warranty coverage. Severity is affected by the change in technology, change in mix, change in labor rates, availability of parts and again the service departments' willingness to use the warranty product. Both internal and external sources of data should be used to finally select a trend factor. Exhibit B shows an internal measure by component for frequency and severity, as well as an external measure of change in severity, using the government's PPI index as a source. For the external measure, we have examined the PPI for auto parts, both new and rebuilt, and for labor charges. We have weighted these indices together to get a combined external index. As labor charges usually make up about half of the total repair bill, we have given it a weight of 50%. We have given auto parts new and rebuilt each a weight of 25%, which assumes that half the time new parts are used in the repair job and half the time rebuilt parts are used.

The selection of annual loss trend factors in auto warranty business is not straightforward. We include external indices in our determination as it is often difficult to explain why internal factors change. For instance, in Exhibit B we show a change in frequency for the new VSC group. This is counterintuitive as it is generally accepted that the quality of new vehicles has improved; shouldn't we then see a decrease in frequency? Perhaps our mix of vehicle make and model has changed. Let's say the we determine that our mix did change. Would we expect the same mix change in the next policy year for which we are projecting rates?

Another problem arises because of the multiyear policy terms. On the new and near-new groups we must wait several years before we become comfortable with projecting a true frequency and severity. We then must use a four or five year old trend factor to project lost costs for the upcoming policy year. We have current calendar year data, but that is a mix of claims from up to seven policy years. If the volume and mix of business is stable over the ratemaking experience, then calendar year trends can be useful, otherwise it can lead to distortions.

It is therefore necessary to include external factors to smooth the results of our internal trend analysis. It is appropriate to give a higher weight to the external factors as they are determined from an industrywide database. This is important because a SCRIP program will most likely get a spread of business from all makes and models. These industrywide or government indices are also important as they tend to smooth the results from internal analysis. As we are often projecting many older policy years in calculating the rate level indication, we must be conscious of the compounding effect of many years of trend to this calculation.

OLDER YEARS: CAN THEY BE USED IN RATEMAKING?

As is seen in Exhibit A, nine policy years have been used in the ratemaking study. We also know from the discussion above that there have been changes over that period of time, most notably the change in the underlying manufacturer's warranty from 1 year / 12,000 miles to 3 years / 36,000 miles. This shift would have a significant impact on the older years. Can these older years be used?

If the TPA or insurer keeps very detailed claim data, an actuary can "as if" the older years. Claims from those older years can actually be recast as if the new terms and conditions were in place. This is helpful not only in getting more accurate projection data but also in calculating loss development factors. Thus older years not only can be used but they are very valuable as they represent truly mature loss data.

IMPORTANCE OF RATEMAKING

The accuracy of the extended warranty rate level indication cannot be stressed enough. Remember, rates are being set on contracts that could be up to seven years in duration. These contracts are a single premium and are non-cancelable by the insurer. Oftentimes it is several years before the adequacy of the current rates can be ascertained, which means you may have written several years of inadequately priced business. If you lower the rates you will most likely lose business and thus revenue just when the claim activity is increasing. It is therefore very important to perform rate level analyses every twelve to eighteen months and make adjustments as necessary.

DEALER MANAGEMENT

The actuary, from the pricing analysis, especially the analysis of frequency, can often find some trouble spots. Notice above that both frequency and severity can be affected by the dealers, or more precisely, the dealers' service departments. It is important, therefore, to keep track of the frequency and severity for each dealer. It is a relatively simple matter to set up a test of significance for an individual dealer's frequency and severity. If either measure is significant, i.e., it is outside the normal range of frequency or severity, than appropriate dealer rehabilitation measures must be taken. By rehabilitation it is meant that the dealer must be put on a program in which frequency and severity are closely monitored, with special reporting done monthly. If within a prescribed time period the dealer's experience has not improved, then the SCRIP will most likely be cancelled. Of course, the TPA (and the insurer) are still responsible for the run-off of the inforce VSCs, which may last up to seven years.

As in any line of insurance, fraud must be guarded against. In the warranty business, you must be vigilant against increases in frequency because severity cannot be changed too drastically. A good dealer management program is a must in this business and the pricing actuary can certainly play an important role.

WHAT TO LOOK OUT FOR:

THE ONLY CONSTANT IS CHANGE

The vehicle service contract ("VSC") industry is young relative to most standard casualty lines of business. As such, it is still evolving. The programs offered by the various third-party administrators of VSCs are constantly changing. These changes in coverage terms and conditions, coverage term options, deductibles and eligibility guidelines are driven by two sets of factors: marketing requirements and changes in the environment of the marketplace. It is important to understand the dynamics of these evolutionary changes and to incorporate such understanding into the ratemaking process.

MARKETING REQUIREMENTS

Innovation is an important marketing tool in the VSC industry. A VSC administrator's need to take an offensive position, to capture or retain market share, generally results in program changes that increase risk. Most VSC administrators rely on a network of independent general agents to distribute their programs to their first-level customers, automobile dealers. Participating auto dealers employ after-sale specialists, finance and insurance ("F&I") managers, to sell VSCs to the second-level customers, automobile purchasers. All auto dealers sell VSCs.

A reasonably effective F&I manager will place a VSC on 30-40% of the retail sales transactions at the dealership. The average profit generated by a VSC sale can add 50-100% to the profit generated by the sale of the vehicle itself. Competition for the auto dealer's business is fierce. Any innovation gives the agent new ammunition to improve his sales pitch. The latest change might have enough impact to tip the account his way. Changes to VSC programs which expand vehicle or mileage eligibility can increase penetration rates at existing accounts. Expanded coverages or benefits give the F&I manager more reasons to justify higher retail pricing, increasing gross profit margins.

ENVIRONMENTAL CHANGES

In opposition to the pro-active nature of marketing-driven changes, environment-driven changes are reactive in nature. The impetus to these changes can come from many directions. Changes in the length or extent of an automobile manufacturer's warranty coverages can require changes in terms, coverages or vehicle rating. Innovations in parts or systems, especially high tech, electronic replacements for existing systems can require changes in coverages, exclusions and rating. Changes in vehicle purchasing patterns can change the makeup of an entire book of business. Ten years ago, a one, two or three year-old vehicle was the hardest used car to sell. Four years ago, such vehicles made up only 10% of VSC sales. Today, such cars account for more than 30% of the VSCs sold.

KNOW WHAT YOU ARE MEASURING

All of the foregoing is meant to illustrate one point. In order to ensure accuracy in ratemaking, especially when measuring trend, know the history of the block of business you are observing. In your due diligence study, prior to starting any rate adequacy study, pay special attention to the following:

Data Integrity – Have all data items, especially manually-coded indicators, been entered and maintained in a consistent manner throughout the history of the database? Are changes in coverage reflected in changes in plan/coverage codes? Run comparison tests on contracts and claims involving similar vehicles/repairs over multiple policy and accident years.

Vehicle Coverage – What changes in coverage options, term/mileage availability, deductible options, vehicle or mileage/age eligibility categories have taken place over the years? When were such changes introduced? Obtain copies of all contracts sold and highlight changes or additions.

Benefits - Have ancillary benefit packages (substitute transportation, towing, trip interruption) changed in composition or in the extent/nature of the benefits provided? Include benefit packages in your comparison of coverage, conditions, exclusions.

Claims Adjustment Policies – What changes have been made in the interpretations of coverages, conditions and exclusions over the years? When were such changes introduced? Obtain copies of all procedure manuals, both external and internal, as well as any pertinent policy memoranda.

Rate Structures – Have there been changes instituted in the method of rating vehicles? Have surcharges been added/dropped? Have vehicle classifications changed? Obtain copies of all rate charts and state premium filing exhibits.

Vehicle Mix – Has the mix of makes, models, equipment changed enough to affect trends in composite loss development patterns or ultimate losses? Has the geographic mix of business changed over the years under study? Obtain historical state/agent loss ratio reports.

IN CONCLUSION

Assessing the impact of change, and the rating provisions employed to offset change, is an essential ingredient in the vehicle service contract ratemaking process. By initially focussing your attention on this aspect of the ratemaking process, you will learn how to apply your analytical skills and techniques to the best advantage.

RATEMAKING WITHOUT HISTORICAL DATA

The most accurate ratemaking is done when there is credible historic program data with which to work. Many times, however, historic data is not available. It may be that the program is new. Oftentimes the program is immature; remember that extended warranty contracts are usually multiyear terms, thus it is usually a number of years before the first policy year is completely expired. It is in these situations that one must use a "back to basics" approach. To price a program properly, one must start with an accurate pure premium, which is the product of frequency times severity.

An interesting example of using a pure premium approach is the pricing of a new program such as the second generation of wind turbines. In the early 1980's, the U.S. government, in an effort to decrease our dependency on foreign oil, granted tax credits for the advancement of alternative energy sources. As part of this initiative, a number of wind turbines were hastily developed and deployed. Each of these machines had manufacturer's warranties, most of which were subsequently insured. Coverage included both mechanical problems and business interruption. Through the ensuing years, the wind turbines proved mechanically deficient and large losses were paid out by insurance companies.

In the mid-nineties, a second generation of wind turbines were being developed and coverage sought for manufacturer's warranties. As there had been problems in the past, the financial backers of these new wind turbines were asking for four specific warranties from the manufacturer; workmanship, efficacy, availability and design defect. Each of these coverages is described in more detail below.

Workmanship – This covers both mechanical breakdown of the machine and the installation of the machine, and would usually be limited to one year from start-up.

Efficacy - This would cover the buyer of the wind turbine for lost revenues as a result of the machine not reaching the promised power generation levels.

Availability – Coverage is given for lost revenues due to down-time in excess of a prescribed number of hours. Total hours functioning would be determined by average sustained wind speed at the field site.

Design Defect – This would cover the retrofitting and lost income due to failure of the wind turbines to perform due to faulty component design. Failure rate thresholds for various components would be established.

Each of the above coverages poses a challenge to the actuary with respect to developing frequency and severity. A thorough examination of the engineering of the new machine must be done. As the actuary is not usually suited for this role, an independent engineering analysis must be sought.

The U.S. and other governments often can provide data on failure rates of similar components (gears, generators, bearings, etc) used in the wind turbine. Deductibles must be established so this does not become a maintenance program and aggregates must also be in place so that a worst-case loss can be determined. Also, as variation exists about all expected values for variables such as failure rates, a risk premium must be considered.

Exhibit C shows a possible approach to determining the pure premium for the above coverages for year 1 of a multiyear manufacturer's warranty. Three separate calculations are made; revenue loss exposure per wind turbine, design defect loss exposure, and materials and workmanship loss exposure.

Revenue loss exposure/wind turbine – This calculation includes the business interruption coverage from both the efficacy and availability sections above. Potential downtimes are given for repairs or retrofits of various components along with the probability that failure of that specific component will occur. For example, given that downtime projected for normal maintenance is 274 hours annually and that 125% of those hours will be used, we can expect 342.5 hours to be used annually in normal maintenance. In total, we expect 1,396.1 hours of downtime; in this program we are allowed 10%, or 876, hours of downtime annually (876 hrs = 10% of 24 hr/day x 365 days) . This is shown at the bottom of Exhibit C, and is the deductible feature of the program. As noted above, about 40% of the deductible would be used for normal maintenance; the other 60% would be to reduce dollar-swapping as well as have the insured share in some risk. With a machine expected to produce 82 kwh/hour, and at $.08/kwh, a resultant loss of $3,412 is expected. A worst-case scenario is also provided, with the probability of occurrence increased by two standard deviations of the expected probability of failure.

Design defect loss exposure – This calculation includes the retrofit cost (severity) and the probability of failure (frequency) by component. Expected costs for each component are calculated; the expected cost per wind turbine for this coverage would be $1,040. The worst-case scenario include revised retrofit costs as well as increased frequencies as described above.

Materials and workmanship loss exposure – As above, a retrofit cost and probability of failure is assigned for each component resulting in an expected cost of failure for each component. The total expected cost for this coverage would be $544. Worst-case scenario is calculated as described above.

As mentioned above, consideration must be given to adding a risk premium to the above. A number of assumptions have been made which, if wrong, can materially affect the calculated pure premium. For instance, the wind turbine is expected to produce 82 kwh per hour. This has not been proven. Also, a rate of $.08 per kwh produced may vary widely in today's fluctuating energy market. Probabilities of failure for similar components tested in government studies might not be representative of the actual components used in the design and manufacture of the wind turbines. In place of a risk premium, a retrospective rating policy might be considered. In any event, while a determination of a pure premium can be made, its accuracy is only as good as the assumptions made. There can be a wide range into which the correct premium may fall.

HANDLING OF UNALLOCATED LOSS ADJUSTMENT EXPENSE

Unallocated loss adjustment expense (ULAE) can be defined as that part of loss adjustment expense which covers the creation and maintenance of a claims department, among other things. It has been overlooked in the past and is one of the reasons why entities have turned to insuring the warranty exposure. Consider the warranty product. The pure premium is typically made up of high frequency low severity occurrences, i.e., there are many small losses. Expected losses in this scenario are generally predictable, and in the early days of shorter-term (mostly annual) warranties the manufacturer kept this risk. As both manufacturer's warranties and extended warranties increased in length of policy term, problems were created. Manufacturers or retail outlets which sold warranties went out of business on occasion, leaving the consumer with a worthless warranty, one on which he most likely paid the premium up front.

A warranty is a promise to pay for a covered repair or replacement to a product; if the provider is not around at the end of a five or ten year policy term, that promise goes unfilled. This is one reason that the transfer of this risk by insurance is now so common. However, insurance companies may decide that they no longer want to be in the warranty business or may go out of business themselves, and non-recognition of ULAE costs can lead to financial difficulties in these instances.

Take for example the auto extended warranty provider. Typically a new-car buyer may purchase an extended warranty for up to seven years or one hundred thousand miles, whichever comes first. The warranty insurer gets the full premium at the time of purchase of the car and is now obligated for the full term of the contract. This means that if for whatever reason the insurer leaves the warranty business, some provision for the fulfilling of the warranty promise must be made. The creation and maintenance of a claims department to fulfill this promise falls under the heading of ULAE and is an important consideration for the actuary in pricing the warranty risk.

Exhibit D illustrates the calculation of ULAE by showing the cost of maintaining a claims operation for the duration of the inforce policies. The calculation starts with the number of claims expected annually, and then the determination of how many underwriters, claim adjusters, auditors and clerks would be needed to service those claims. Also factored in would be the cost of equipment, and facilities for these people. As can be seen, the total cost can then be reduced to a rate per contract and included in total price.

The most important calculation in Exhibit D may be the of distribution of claims. In this example warranty contracts are sold with terms varying from 1 year to 7 years. For policy year 1998, the contract sold on December 31st of that year will not expire until December 31, 2005. If no more contracts were ever written, there would be a need for a claims staff for seven more years. It is important that claims data can be linked to policy information in order to determine the claim development (it is not uncommon for warranty administrators to keep premium and claims data completely separated, though this is becoming less and less common). If no data is available, a distribution can be developed by working with sources knowledgeable with the product being warrantied. There may also be similar products being warrantied about which claim development data is available that can be used as a proxy.

The actual ULAE costs can be determined in one of two ways. The costs may be determined by viewing the claims operation either as an on-going business or as a run-off operation. Viewing it as a run-off operation would lower the costs as claim-paying standards would most likely drop. The insurer is no longer interested in maintaining a strong service image. For example, in an on-going operation the standard of issuing a claim payment from notice of claim may be five days; in a run-off operation this standard could be relaxed to two weeks or more. This philosophy would also influence the setting of a ULAE reserve.

ULAE can be collected in various ways, depending on the way an insurer provides the warranty product. If the insurer administers the settling of the claims it can be included in the warranty premium. If a third party administrator (TPA) handles the claims, it may be provided for by fees charged by the administrator to the dealer or retailer or it may be part of the commission structure. For example, the TPA may earn a commission of 25%, but only get 15% with the remaining 10% amortized over several years.

As can be seen, not recognizing the ULAE costs on a multiyear non-cancelable policy can have financial implications. At the very least, a liability should be shown in the financial statements. At the worst, it could lead to a claims department totally unprepared to handle the volume of claims in the future.

PITFALLS

Many companies have entered the insuring of manufacturers' and extended warranty market and many have failed, losing great amounts of money. Most often failures occur because the risk being transferred was not understood. Let's face it, at the outset, this business looks very attractive, as for the most part premiums are paid up front in full, and claims may occur years later. Just think of all the investment income to be made!

Vehicle Service Contracts

Vehicle service contracts ("VSCs") present us with a unique risk/exposure structure. In no other form of insurance is the insured, the producing agent and the service provider the same entity. This structure is akin to a doctor selling health insurance to his own patients. As you might imagine, such a structure is full of moral hazards and conflicts of interest.

As a companion function to careful ratemaking, account management is a necessity to ensure the success of any VSC program. Opportunities abound for unscrupulous dealership personnel to take advantage of a VSC program. Used car sales managers can increase gross profits by avoiding reconditioning expenses and having failed vehicles repaired under the VSC program. F&I managers can increase gross profits by posting incorrect issue mileages in order to reduce premiums below required levels, while maintaining high retail rates (retail rates are only controlled in Florida). Service departments can comb over each car in order to "discover" claims.

None of the previous examples can be controlled through underwriting or claims adjustment efforts or controls. Without effective account management systems, administrators are left with three, equally unpalatable alternatives: raise rates, post-claims underwrite or cancel bad accounts. If rates are raised beyond competitive levels, business will fall off. Generally, the greatest losses are among the lowest risk, most profitable vehicle makes. The artificially high rates become attractive only to high-risk dealers, selling high-risk cars, which will soon prove even the artificially high rates to be inadequate. Tightening claim adjustment policies can have the same effect – lost business. Cancellation of poorly-performing accounts, while eliminating the problem, can end up eliminating all of a company's problems.

Information flow is the cornerstone of a successful account management system. Situations can change quickly in the automobile business. A monthly exception report, listing and classifying all poorly-performing accounts, is absolutely necessary. Also necessary is an experienced, well trained staff to manage the recovery process. The overall concept of account management is to identify problem accounts, to identify the specific problem areas within the operation of such accounts and then to take corrective action.

Identifying problem accounts is simply a matter of generating a listing of accounts whose earned loss ratios exceed a specific target. The three major areas of VSC groupings involve new vehicles, near-new vehicles (or extended eligibility new vehicles) and used vehicles. If any or all of the target loss ratios for these groupings are exceeded, the account should show on the listing. If programming resources permit, it is also useful to develop some sort of ranking system, encompassing factors such as : newly acquired account shock losses, number of VSC grouping target loss ratios exceeded, overall loss ratio target loss ratio exceed, as well as the amount by which the targets have been exceeded.

Identifying problem areas within the operation of the targeted accounts is a more complex issue. In order to begin the analysis of specific problem areas, a more complex target set, or model, is necessary. This model needs to be constructed according to major franchise group (Standard Asian, luxury Asian, standard domestic, luxury domestic, standard European, luxury European) and reflect acceptable frequency and severity targets for each VSC grouping (New, Near-New, Used, Total). Frequency and severity targets for this matrix can be calculated by averaging the results of several accounts within each franchise group whose loss ratios for all VSC groupings are at or below target levels.

Once the variances from frequency/severity targets are established, specific causes for such variances can be derived and solutions proposed. High rates of early used vehicle claims can be traced to less than adequate used vehicle reconditioning practices. Generally high claim severity (usually combined with high rates of multi-item repairs) usually point to highly incentivised service writers/technicians "discovering" failures that were not prompted by customer complaints. Generally, high frequency levels point to some type of customer incentive program, e.g. free inspections or other service specials.

In order to implement solutions, the internal systems and the state filings must be flexible enough to provide support for: reduced claim reimbursement (factory time and/or labor rates as opposed to retail) claim elimination periods (typically 30 days on used vehicles) premium adjustments (individual rate premium modifier factors) underwriting restrictions (high mileage used vehicles, long term new vehicle plans). Rate adjustments, elimination periods and underwriting restrictions are used to address selection and reconditioning issues, involving the sale of the VSCs. Reduced claim reimbursement is used to combat overzealousness in the service department. By focussing the solutions on the specific areas of the account's operation that is causing the problem, recovery is speeded and recovery rates are increased.

WARRANTY IN GENERAL

In the early days companies evaluated warranty business on a calendar year basis. Premiums on multi-year terms were earned evenly over the contract period. Unfortunately, losses tended to occur later in the term of the warranty. In Exhibit E, it can be seen how this combination understates the loss ratio in the first calendar year of the warranty term. Now, since the loss ratio is so low, an obvious albeit erroneous conclusion would be that not only should we write more of this business, we should reduce rates to help our marketers! It only takes a few years to dig a deep hole, as inadequately priced business has now been written for several years. Rate relief is essential. Of course, this leads to further problems. If the rate level increase needed is large, there may be difficulty getting approval from the various states. Even if approvals are finally received, implementing a large rate increase could lead to a very rapid drop-off in VSCs written, as dealers can use a competing program. A large drop-off in VSCs would mean a large reduction in revenue, just when the cash is needed to pay the claims from the old business. It is easy to see how this could become a run-off operation.

Earning premiums correctly is very important as can be seen above. Premiums should be earned in direct proportion to the loss payout pattern. Earning premiums in this fashion maintains the proper loss ratio for the life of the policy period, as shown in Exhibit E. Hopefully existing loss payout data is available in order to determine the payout pattern. In cases where the data is not available and the losses are expected to start out slowly in the beginning of the term and monotonically increase over the life of the contracts, the reverse sum of the digits rule can be used. Exhibit E shows the loss payout pattern described by this rule. As shown, we would earn 1/36 of the premium in the first year, 2/36 in the second year, and so on up to 8/36 in the last year. Note that this earning methodology is conservative; it does not recognize the aforementioned "attrition factor." The state of Louisiana actually requires that a non-insurance company that guarantees warranties or extended warranties earn its income no faster than the reverse sum of the digits rule. If the term of the contract period is annual, this rule is often referred to as the reverse rule of 78s (using monthly earnings).

Pricing of warranties or extended warranties should be done by product or at most by homogeneous classes of products. Do not make the mistake of giving one overall rate for a warranty program made up of many different products. Exhibit F, example 1, illustrates what can happen. Company A administers a warranty program for "brown and white goods" (basically electronics appliances, and office equipment.) Loss costs are available, and Company A is looking to transfer the warranty risk to insurance company B. Since B will insure the entire program, B decides to give a single program rate of $169. Unfortunately for B, A writes a new account which only sells refrigerators. This changes the mix of risks, thus changing loss costs and making the single rate of $169 inadequate as the new rate should be $193. Practically speaking, rates would not be modified every time a new account came on line, so it would be better to charge a rate by class to minimize the mix change problem.

Another pricing pitfall to avoid is basing the rate on the overall revenue an administrator gets for the warranty contract. Again in Exhibit F, example 2 Company A (the administrator) sells a warranty contract for $50, and Company B (the insurer) determines that the loss cost is $5. B than grosses the loss cost up for expenses and wants $7 in premium. B than sets a rate of 14% per revenue. Unfortunately for B, next year A decides to lower its selling price of the warranty to $40. Now B only gets $5.60 per contract, which barely covers his loss costs let alone his expenses.

Another problem often encountered by the pricing actuary on warranty business is the lack of quality data. To properly price a warranty product, policy year data must be used. Most administrators do not show data in policy year format; some cannot show it as losses cannot be tied back to the premium. Obviously in this type of operation there can be no verification of coverage; the claim is paid when it is presented. This type of account cannot be soundly priced. If triangular data is available, it must be reconciled with the TPA's audited financials. Again, many TPAs are not used to providing actuarial data, so a thorough checking of the data is required.

CONCLUSION

Ratemaking in the extended warranty line of business is well-suited to take advantage of the actuarial approach. The business is driven by frequency rather than severity so that it lends itself to actuarial modeling. For the vehicle extended warranty there is often credible data available. When there is not data available, the "back-to-basics" approach is best done by an actuary. The actuary is an essential member of the warranty pricing team.

AUTO EXTENDED WARRANTY **EXHIBIT A**

LOSS PROJECTIONS

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Policy Year	Number of Contracts	Ultimate Written Premium	Average Premium	Paid Losses To Date	Paid Loss Ratio	Paid LDF	IBNR Factor	ELR	Paid Projection	Stanard-Buhlmann	B-F Projection	Pure Premium /Contract	Cumulative Pure Premium	Pure Premium Projection
1990	25,000	$ 5,000,000	$ 200	$ 4,800,000	96.0%	1.000	0.000	85.0%	$ 4,800,000	$ 4,800,000	$ 4,800,000	$ 192	$ 192	4,800,000
1991	25,000	$ 5,050,000	$ 202	$ 5,201,500	103.0%	1.000	0.000	85.0%	$ 5,201,500	$ 5,201,500	$ 5,201,500	$ 208	$ 200	$ 5,000,750
1992	30,000	$ 6,150,000	$ 205	$ 6,211,500	101.0%	1.030	0.029	85.0%	$ 6,397,845	$ 6,377,189	$ 6,363,757	$ 212	$ 205	$ 6,136,971
1993	35,000	$ 7,350,000	$ 210	$ 5,953,500	81.0%	1.140	0.123	85.0%	$ 6,786,990	$ 6,788,419	$ 6,720,737	$ 192	$ 201	$ 7,026,172
1994	40,000	$ 8,520,000	$ 213	$ 5,367,600	63.0%	1.450	0.310	85.0%	$ 7,783,020	$ 7,813,382	$ 7,615,117	$ 190	$ 198	$ 7,922,867
1995	45,000	$ 9,675,000	$ 215	$ 3,676,500	38.0%	2.300	0.565	85.0%	$ 8,455,950	$ 8,734,748	$ 8,324,707	$ 185	$ 195	$ 8,780,809
1996	50,000	$ 11,000,000	$ 220	$ 1,100,000	10.0%	5.400	0.815	85.0%	$ 5,940,000	$ 9,390,586	$ 8,718,519	$ 174	$ 191	$ 9,548,867
1997	55,000	$ 12,375,000	$ 225	$ 111,375	0.9%	27.700	0.964	85.0%	$ 3,085,088	$ 11,144,799	$10,250,387	$ 186	$ 190	$10,458,065
1998	60,000	$ 13,800,000	$ 230	$ 13,800	0.1%	432.000	0.998	85.0%	$ 5,961,600	$ 12,749,013	$11,716,647	$ 195	$ 191	$11,459,403

column	notes on calculation
4	col 3 / col 2
6	col 5 / col 3
8	1.00 - 1.00 / col 7
10	col 5 x col 7
11	====>
12	col 5 + col 3 x col 8 x col 9
13	col 12 / col 2
14	(sum of col 12 current and preceding years) / (sum of col 2 current and preceding years)
15	col 2 x col 14

Stanard / Buhlmann Calculation

	ELR est	prem	1-lag	premium x 1.0 - lag	IBNR	reported losses	ultimate losses
1990	92%	$ 5,000,000	0.000	$ -	0	$ 4,800,000	$ 4,800,000
1991	92%	$ 5,050,000	0.000	$ -	$ -	$ 5,201,500	$ 5,201,500
1992	92%	$ 6,150,000	0.029	$ 179,126	$ 165,689	$ 6,211,500	$ 6,377,189
1993	92%	$ 7,350,000	0.123	$ 902,632	$ 834,919	$ 5,953,500	$ 6,788,419
1994	92%	$ 8,520,000	0.310	$ 2,644,138	$ 2,445,782	$ 5,367,600	$ 7,813,382
1995	92%	$ 9,675,000	0.565	$ 5,468,478	$ 5,058,248	$ 3,676,500	$ 8,734,748
1996	92%	$ 11,000,000	0.815	$ 8,962,963	$ 8,290,586	$ 1,100,000	$ 9,390,586
1997	92%	$ 12,375,000	0.964	$ 11,928,249	$ 11,033,424	$ 111,375	$ 11,144,799
1998	92%	$ 13,800,000	0.998	$ 13,768,056	$ 12,735,213	$ 13,800	$ 12,749,013
totals		$ 78,920,000		$ 43,853,642		$32,435,775	

<u>Calculation of ELR Estimate</u>

sb ibnr = elr est x	$43,853,642			
elr est = [ibnr est +	$32,435,775	] /	$ 78,920,000	
ibnr est = elr est x	$78,920,000	-	$ 32,435,775	
elr est x	$43,853,642	= elr est x	$ 78,920,000	$ 32,435,775
	$32,435,775	= elr est x	$ 35,066,358	
	0.92	= elr est		

AUTO EXTENDED WARRANTY

TREND ANALYSIS

EXHIBIT B

INTERNAL REPAIR COST ANALYSIS

1999

component	coverage	policy count	claim count	payment	frequency	severity
rental	new	75,000	15,000	$ 1,050,000	20.0%	$ 70
water pump	new	75,000	6,000	$ 1,050,000	8.0%	$ 175
air cond. compressor	new	75,000	3,750	$ 1,500,000	5.0%	$ 400
fuel pump	new	75,000	2,250	$ 528,750	3.0%	$ 235
transmission (automatic) internal parts	new	75,000	2,250	$ 1,743,750	3.0%	$ 775
transaxle (automatic) internal parts	new	75,000	2,250	$ 1,912,500	3.0%	$ 850
transaxle (automatic) assembly	new	75,000	1,125	$ 1,293,750	1.5%	$ 1,150
transmission (automatic) assembly	new	75,000	750	$ 750,000	1.0%	$ 1,000
engine assembly	new	75,000	450	$ 675,000	0.6%	$ 1,500
differential (rear) assembly	new	75,000	300	$ 157,500	0.4%	$ 525
transmission (manual) assembly	new	75,000	75	$ 62,250	0.1%	$ 830
subtotal	new	75,000	34,200	$10,723,500	45.6%	$ 314

1998

component	policy count	claim count	payment	frequency	severity
rental	60,000	11,400	$ 855,000	19.0%	$ 75
water pump	60,000	4,200	$ 735,000	7.0%	$ 175
air cond. compressor	60,000	2,700	$ 1,012,500	4.5%	$ 375
fuel pump	60,000	2,100	$ 472,500	3.5%	$ 225
transmission (automatic) internal parts	60,000	1,500	$ 1,155,000	2.5%	$ 770
transaxle (automatic) internal parts	60,000	1,500	$ 1,275,000	2.5%	$ 850
transaxle (automatic) assembly	60,000	900	$ 1,080,000	1.5%	$ 1,200
transmission (automatic) assembly	60,000	600	$ 660,000	1.0%	$ 1,100
engine assembly	60,000	360	$ 522,000	0.6%	$ 1,450
differential (rear) assembly	60,000	180	$ 94,500	0.3%	$ 525
transmission (manual) assembly	60,000	60	$ 66,000	0.1%	$ 1,100
subtotal	60,000	25,500	$ 7,927,500	42.5%	$ 311

component	change in frequency	change in severity	loss trend factor
rental	5.3%	-6.7%	-1.8%
water pump	14.3%	0.0%	14.3%
air cond. compressor	11.1%	6.7%	18.5%
fuel pump	-14.3%	4.4%	-10.5%
transmission (automatic) internal parts	20.0%	0.6%	20.8%
transaxle (automatic) internal parts	20.0%	0.0%	20.0%
transaxle (automatic) assembly	0.0%	-4.2%	-4.2%
transmission (automatic) assembly	0.0%	-9.1%	-9.1%
engine assembly	0.0%	3.4%	3.4%
differential (rear) assembly	33.3%	0.0%	33.3%
transmission (manual) assembly	0.0%	-24.5%	-24.5%
subtotal	7.3%	0.9%	8.2%

EXTERNAL TREND ANALYSIS

Producer Price Index Data

	Annual Trend in Index				
	Parts New	Parts Rebuilt	Labor	Combined Annual Trend	index
weights	0.25	0.25	0.5		1.000
1990	1.1%	0.0%	4.8%	2.7%	1.027
1991	0.6%	0.0%	3.8%	2.1%	1.048
1992	0.6%	2.2%	3.2%	2.3%	1.072
1993	0.4%	0.5%	2.7%	1.6%	1.089
1994	1.5%	1.3%	2.3%	1.9%	1.109
1995	0.2%	-0.4%	2.3%	1.1%	1.121
1996	-0.7%	-1.5%	3.8%	1.4%	1.136
1997	-0.6%	-0.8%	3.5%	1.4%	1.152

exponential trend = 1.8%

WIND TURBINES

EXHIBIT C

YEAR 1 EXPOSURE

revenue loss exposure/wind turbine

item	potential downtime (hrs/yr)	probability of occurrence	probable downtime hours	potential downtime (hrs/yr)	worst case probability	worst case downtime hours
normal maintenance	274	1.25	342.5	274	1.96	536.2
rotor blade repair	1440	0.08	115.2	1440	0.22	318.8
hub retrofit	1800	0.10	180.0	1800	0.52	943.6
teeter damper retrofit	96	0.15	14.4	96	0.77	74.1
gearbox retrofit	2160	0.15	324.0	2160	0.49	1057.0
generator retrofit	2160	0.10	216.0	2160	0.24	521.4
mainframe repair	96	0.15	14.4	96	0.49	47.0
yaw bearing retrofit	1800	0.10	180.0	1800	0.27	485.4
tower repair	96	0.10	9.6	96	0.58	55.8
totals			1396.1			4039.3
revenue loss @$0.08/kwh; 90% availability 876 hours allowable projected 82 kwh/hour			$3,412			$20,751

Design Defect loss exposure

item	retrofit cost	probability of occurrence	probable cost exposure	reasonable retrofit cost	worst case probability occurrence	worst case exposure
normal maintenance	n/a	1.00	$0	n/a	2.00	$0
rotor blade repair	$2,000	0.05	$100	$16,500	0.08	$1,292
hub retrofit	$5,000	0.06	$300	$6,500	0.30	$1,952
teeter damper retrofit	$2,500	0.08	$200	$4,000	0.45	$1,791
gearbox retrofit	$4,000	0.06	$240	$19,200	0.16	$3,052
generator retrofit	$1,000	0.04	$40	$7,000	0.07	$478
mainframe repair	$0	0.06	$0	$7,500	0.16	$1,192
yaw bearing retrofit	$4,000	0.04	$160	$5,500	0.15	$842
tower repair	$0	0.08	$0	$3,000	0.31	$919
totals			$1,040			$11,518

materials and workmanship loss exposure

item	retrofit cost	probability of occurrence	probable cost exposure	reasonable retrofit cost	worst case probability occurrence	worst case exposure
normal maintenance	n/a	n/a	$0	n/a	2.00	$0
rotor blade repair	$2,000	0.04	$80	$2,000	0.15	$306
hub retrofit	$2,000	0.04	$80	$2,000	0.22	$448
teeter damper retrofit	$700	0.04	$28	$700	0.29	$206
gearbox retrofit	$3,500	0.06	$210	$3,500	0.30	$1,051
generator retrofit	$1,200	0.04	$48	$1,200	0.15	$184
mainframe repair	$700	0.06	$42	$700	0.30	$210
yaw bearing retrofit	$2,800	0.02	$56	$2,800	0.08	$214
tower repair	$0	0.04	$0	$0	0.29	$0
totals			$544			$2,620

Oustanding ULAE Estimates
As of 10/97

Assumptions

Contract, Claims Development, & Frequency

	PY1991	PY1992	PY1993	PY1994	PY1995	PY1996	PY1997	PY1998
Number of Written Contracts	**85,000**	**90,000**	**95,000**	**100,000**	**105,000**	**110,000**	**115,000**	**120,000**
Processing Frequency Ratio	**110%**	110%	110%	110%	110%	110%	110%	110%
Projected Processed Claims	93,500	99,000	104,500	110,000	115,500	121,000	126,500	132,000

	1yr	2yr	3yr	4yr	5yr	6yr	7yr	8yr
Cumulative Claim Development Pattern	**25%**	**40%**	**55%**	**75%**	**85%**	**95%**	**98%**	**100%**
Incremental Claim Development Pattern	25%	15%	15%	20%	10%	10%	3%	2%

Projected Processed Claims

	CY1998	CY1999	CY2000	CY2001	CY2002	CY2003	CY2004	CY2005	Total O/S
PY98	33,000	19,800	19,800	26,400	13,200	13,200	3,960	2,640	132,000
PY97	18,975	18,975	25,300	12,650	12,650	3,795	2,530		94,875
PY96	18,150	24,200	12,100	12,100	3,630	2,420			72,600
PY95	23,100	11,550	11,550	3,465	2,310				51,975
PY94	11,000	11,000	3,300	2,200					27,500
PY93	10,450	3,135	2,090						15,675
PY92	2,970	1,980							4,950
PY91	1,870								1,870
PY91-97	86,515	70,840	54,340	30,415	18,590	6,215	2,530		269,445
PY98	33,000	19,800	19,800	26,400	13,200	13,200	3,960	2,640	132,000

Personnel & Plant Costs

	Adjuster	Underwriter	Auditor	Clerk
Average Salary	**40,000**	**45,000**	**50,000**	**25,000**
Avg No of Claims Processed per Day	**20**	**110**	**325**	**110**
Avg No of Claims Processed per Yr	**5,000**	**27,500**	**75,000**	**27,500**
Benefits(as %age of salary)	**17%**	**17%**	**17%**	**17%**
Rent per Square Foot	**15**	**15**	**15**	**15**
Square Feet Per Person	**200**	**200**	**200**	**200**
Total Rent Cost per Person	3,000	3,000	3,000	3,000
Equipment	**315**	**315**	**315**	**315**
Telephone per Claim Cost	**4.50**	**4.50**	**4.50**	**4.50**
Misc per Claim Costs(postage, electric, etc.)	**1.00**	**1.00**	**1.00**	**1.00**

Scenario Options

☑ Round Up Required Personnel

☑ Run off Mode

Personnel Effeciency Level

94.0%

Plant Cost Savings

0.0%

Inflation Factors

Annual Wage Inflation Rate	**5.0%**
Annual Plant Cost Inflation Rate	**2.0%**

Dates

Run Off Date
12/31/1997

Summary of Results

	Total	Avg/Clm	Avg/Pol
ULAE Estimate for PY91-97	**6,746,891**	**25.0**	**27.54**
ULAE Estimate for PY98	**2,956,620**	**22.4**	**24.64**

ULAE Projections

Prior Book

PY91-97	CY	Expected Claim Counts	Required Number of Personnel: Adjusters	Underwriters	Auditors	Clerks	Annual Personnel Costs: Adjusters	Underwriters	Auditors	Clerks	Benefits	Tot Wages & Benefits	Wage Inflation	Total
	1998	86,515	19.0	4.0	2.0	4.0	760,000	180,000	100,000	100,000	193,800	1,333,800	5.0%	1,333,800
	1999	70,840	16.0	3.0	1.0	3.0	640,000	135,000	50,000	75,000	153,000	1,053,000	5.0%	1,105,650
	2000	54,340	12.0	2.0	1.0	2.0	480,000	90,000	50,000	50,000	113,900	783,900	5.0%	864,250
	2001	30,415	7.0	2.0	1.0	2.0	280,000	90,000	50,000	50,000	79,900	549,900	5.0%	636,578
	2002	18,590	4.0	1.0	1.0	1.0	160,000	45,000	50,000	25,000	47,600	327,600	5.0%	398,200
	2003	6,215	2.0	1.0	1.0	1.0	80,000	45,000	50,000	25,000	34,000	234,000	5.0%	298,650
	2004	2,530	1.0	1.0	1.0	1.0	40,000	45,000	50,000	25,000	27,200	187,200	5.0%	250,866
	Totals	269,445	61	14	8	14	2,440,000	630,000	400,000	350,000	649,400	4,469,400		4,887,993

CY	Plant Costs: Telephone	Rent	Equipment	Other	Total	Plant Inflation	Total Plant	Total ULAE Estimate	Avg ULAE Per Claim
1998	389,318	87,000	9,135	86,515	571,968	2.0%	571,968	1,905,768	22
1999	318,780	69,000	7,245	70,840	465,865	2.0%	475,182	1,580,832	22
2000	244,530	51,000	5,355	54,340	355,225	2.0%	369,576	1,233,826	23
2001	136,868	36,000	3,780	30,415	207,063	2.0%	219,736	856,314	28
2002	83,655	21,000	2,205	18,590	125,450	2.0%	135,791	533,991	29
2003	27,968	15,000	1,575	6,215	50,758	2.0%	56,040	354,690	57
2004	11,385	12,000	1,260	2,530	27,175	2.0%	30,603	281,469	111
Totals	1,212,503	291,000	30,555	269,445	1,803,503		1,858,897	**6,746,891**	**25**

EXHIBIT D
PAGE 3

New Book

PY98	CY	Expected Claim Counts	Required Number of Personnel: Adjusters	Underwriters	Auditors	Clerks	Annual Personnel Costs: Adjusters	Underwriters	Auditors	Clerks	Benefits	Tot Wages & Benefits	Wage Inflation	Total
	1998	33,000	7.0	2.0	1.0	2.0	280,000	90,000	50,000	50,000	79,900	549,900	5.0%	**549,900**
	1999	19,800	4.0	1.0	1.0	1.0	160,000	45,000	50,000	25,000	47,600	327,600	5.0%	**343,980**
	2000	19,800	4.0	1.0	1.0	1.0	160,000	45,000	50,000	25,000	47,600	327,600	5.0%	**361,179**
	2001	26,400	6.0	1.0	1.0	1.0	240,000	45,000	50,000	25,000	61,200	421,200	5.0%	**487,592**
	2002	13,200	3.0	1.0	1.0	1.0	120,000	45,000	50,000	25,000	40,800	280,800	5.0%	**341,314**
	2003	13,200	3.0	1.0	1.0	1.0	120,000	45,000	50,000	25,000	40,800	280,800	5.0%	**358,380**
	2004	3,960	1.0	1.0	1.0	1.0	40,000	45,000	50,000	25,000	27,200	187,200	5.0%	**250,866**
	2005	2,640	1.0	1.0	1.0	1.0	40,000	45,000	50,000	25,000	27,200	187,200	5.0%	**263,409**
Totals		132,000	29	9	8	9	1,160,000	405,000	400,000	225,000	372,300	2,562,300		**2,956,620**

	CY	Plant Costs: Telephone	Rent	Equipment	Other	Total	Plant Inflation	Total Plant	Total ULAE Estimate	Avg ULAE Per Claim
	1998	148,500	36,000	3,780	33,000	-	2.0%	-	549,900	17
	1999	89,100	21,000	2,205	19,800	-	2.0%	-	343,980	17
	2000	89,100	21,000	2,205	19,800	-	2.0%	-	361,179	18
	2001	118,800	27,000	2,835	26,400	-	2.0%	-	487,592	18
	2002	59,400	18,000	1,890	13,200	-	2.0%	-	341,314	26
	2003	59,400	18,000	1,890	13,200	-	2.0%	-	358,380	27
	2004	17,820	12,000	1,260	3,960	-	2.0%	-	250,866	63
	2005	11,880	12,000	1,260	2,640	-	2.0%	-	263,409	100
Totals		594,000	165,000	17,325	132,000	-		-	**2,956,620**	**22**

EXTENDED WARRANTY — EXHIBIT E

EARNING OF PREMIUM IN POLICY YEAR X

LOSS PAYOUT PATTERN BY CALENDAR YEAR

X	X + 1	X + 2	X + 3	X + 4	X + 5	X + 6	X + 7	TOTAL
3%	7%	10%	12%	15%	20%	20%	13%	100%

For use with examples below:

POLICY YEAR WRITTEN PREMIUM = $ 100,000

EXPECTED LOSS RATIO = 75%

EXPECTED LOSSES = $ 75,000

EXAMPLE 1: PREMIUMS EARNED PRO-RATA

	X	X + 1	X + 2	X + 3	X + 4	X + 5	X + 6	X + 7	TOTAL
Earned Premium	$ 12,500	$ 12,500	$ 12,500	$ 12,500	$ 12,500	$ 12,500	$ 12,500	$ 12,500	$ 100,000
Cumulative Earnings	$ 12,500	$ 25,000	$ 37,500	$ 50,000	$ 62,500	$ 75,000	$ 87,500	$100,000	
Incurred Losses	$ 2,250	$ 5,250	$ 7,500	$ 9,000	$ 11,250	$ 15,000	$ 15,000	$ 9,750	$ 75,000
Cumulative Losses	$ 2,250	$ 7,500	$ 15,000	$ 24,000	$ 35,250	$ 50,250	$ 65,250	$ 75,000	
Policy Year X Loss Ratio	18%	30%	40%	48%	56%	67%	75%	75%	

EXAMPLE 2: PREMIUMS EARNED IN PROPORTION TO LOSS PAYOUT PATTERN

	X	X + 1	X + 2	X + 3	X + 4	X + 5	X + 6	X + 7	TOTAL
Earned Premium	$ 3,000	$ 7,000	$ 10,000	$ 12,000	$ 15,000	$ 20,000	$ 20,000	$ 13,000	$ 100,000
Cumulative Earnings	$ 3,000	$ 10,000	$ 20,000	$ 32,000	$ 47,000	$ 67,000	$ 87,000	$100,000	
Incurred Losses	$ 2,250	$ 5,250	$ 7,500	$ 9,000	$ 11,250	$ 15,000	$ 15,000	$ 9,750	$ 75,000
Cumulative Losses	$ 2,250	$ 7,500	$ 15,000	$ 24,000	$ 35,250	$ 50,250	$ 65,250	$ 75,000	
Policy Year X Loss Ratio	75%	75%	75%	75%	75%	75%	75%	75%	

EXAMPLE 3: PREMIUMS EARNED IN PROPORTION TO REVERSE SUM OF THE DIGIT RULE

Earnings done over 8 years, thus sum of digits = (n)(n+1)/2 = 8 x 9 / 2 = 36

	X	X + 1	X + 2	X + 3	X + 4	X + 5	X + 6	X + 7	TOTAL
Earned Premium Pattern	1/36	2/36	3/36	4/36	5/36	6/36	7/36	8/36	36/36
Earned Premiums	$ 2,778	$ 5,556	$ 8,333	$ 11,111	$ 13,889	$ 16,667	$ 19,444	$ 22,222	$ 100,000
Cumulative Earnings	$ 2,778	$ 8,333	$ 16,667	$ 27,778	$ 41,667	$ 58,333	$ 77,778	$100,000	
Incurred Losses	$ 2,250	$ 5,250	$ 7,500	$ 9,000	$ 11,250	$ 15,000	$ 15,000	$ 9,750	$ 75,000
Cumulative Losses	$ 2,250	$ 7,500	$ 15,000	$ 24,000	$ 35,250	$ 50,250	$ 65,250	$ 75,000	
Policy Year X Loss Ratio	81%	90%	90%	86%	85%	86%	84%	75%	

EXTENDED WARRANTY **EXHIBIT F**

RATING PROBLEM CAUSED BY CHANGE IN MIX

EXAMPLE 1: CHANGE IN LOSS COSTS DUE TO CHANGE IN MIX - TPA ADDS REFRIGERATOR ACCOUNT

ORIGINAL MIX OF RISKS

	number of contracts	loss costs per contract	total loss costs
VCRs	200	$ 80	$ 16,000
Refrigerators	200	$ 200	$ 40,000
copiers	200	$ 100	$ 20,000
totals	600	$ 127	$ 76,000

CALCULATION OF SINGLE RATE

Loss cost per contract	=	$ 127
Expected loss ratio	=	75%
Gross rate per contract	=	$ 169

NEW MIX OF RISKS

	number of contracts	loss costs per contract	total loss costs
VCRs	200	$ 80	$ 16,000
Refrigerators	400	$ 200	$ 80,000
copiers	200	$ 100	$ 20,000
totals	800	$ 145	$ 116,000

CALCULATION OF SINGLE RATE

Loss cost per contract	=	$ 145
Expected loss ratio	=	75%
Gross rate per contract	=	$ 193

EXAMPLE 2: GROSS RATE BASED ON REVENUE

	Price of Warranty	Loss Costs	Insurer Expenses	Insurance Costs	IC as % of Price		Insurance Premium
Year 1	$ 50.00	$ 5.00	$ 2.00	$ 7.00	14%		$ 7.00
Year 2 - rate still 14% of price	$ 40.00	$ 5.00	$ 2.00	$ 7.00	18%		$ 5.60

216

A Macro Validation Dataset for U.S. Hurricane Models

Douglas J. Collins, FCAS, MAAA and
Stephen P. Lowe, FCAS, MAAA

A Macro Validation Dataset for U.S. Hurricane Models

By Douglas J. Collins and Stephen P. Lowe

Abstract

Public and regulatory acceptance of catastrophe models has been hampered by the complexity and proprietary nature of the models. The outside user is generally dependent on the modeler to demonstrate the validity and reasonableness of model results. Accordingly, we have developed a dataset permitting macro validation - one that would allow a lay person to compare the overall results of a hurricane model to an historical record.

The macro validation dataset consists of the aggregate insured losses from hurricanes affecting the continental United States from 1900 through 1999. The historical losses in each county have been "trended" – adjusted from the conditions at the time to those existing today. The trending reflects not only estimated changes in price levels, but also estimated changes in the value of the stock of properties and contents, and changes in the insurance system. Our work extends and improves upon similar work by Landsea and Pielke (1998), published by the American Meteorological Society.

The paper describes the construction of the validation dataset and summarizes the resulting size of loss distributions by event, state and county. It also provides tables summarizing key statistics about all hurricanes affecting the United States (and Puerto Rico, the U.S. Virgin Islands and Bermuda) during the 20th century. Finally, we compare summary statistics from the dataset to the results of a hypothetical probabilistic hurricane model.

I. INTRODUCTION

Hurricane Andrew in 1992 heightened the concern among property insurers and reinsurers about the potential for losses from natural catastrophes. This heightened concern spread beyond hurricanes to other perils with the Northridge earthquake in 1994, and several major winter snowstorms and tornadoes during the nineties. Major catastrophes outside the U.S. during this time have also helped keep catastrophe issues in the forefront for property insurers and reinsurers worldwide.

Since natural catastrophes are infrequent, traditional actuarial pricing methods are of limited value. Actuaries are accustomed to estimating rate adequacy by adjusting a body of historical insurance premium and loss experience to reflect the anticipated future environment. For property insurance, this typically involves a projection using three to six years of recent, mature experience. Prior to hurricane Andrew, the actuarial literature suggested using a thirty-year experience period for measuring excess wind loads in property insurance ratemaking.

When extreme events in a particular region are expected to happen only once every hundred years or more, alternative approaches are clearly required. This is true whether the objective is to measure expected losses for rating purposes or probable maximum losses[1] for risk and capital management purposes. For catastrophe risk management, probabilistic computer simulation models have been developed as such an alternative. These models incorporate longer-term historical data about the physical events as well as engineering knowledge about their destructive potential. Insurers, reinsurers and rating agencies have generally accepted use of the models to project losses.

The models and their use as a ratemaking tool have not been free from controversy. Some insurance regulators have rejected their use in rate filings, citing the difficulty of verifying the model results. Regulators have also cited extreme rate indications and inconsistent results between competing models as a basis of their rejection. Despite these issues, the use of models continues to increase because they provide the most comprehensive use of available data to measure the costs and risks of catastrophes. In response, regulators in Florida and Louisiana have set up formal processes for evaluating catastrophe models.

Model Validation

Fundamentally, all catastrophe models proceed along the same analytical path. First, the key scientific parameters describing a specific historical or hypothetical event are determined. The models then estimate the incidence of damaging forces to property from that event. Finally, the resulting property damage and insured loss are

[1] The probable maximum loss, or PML, is the loss amount that is estimated to be exceeded with a specific probability, for example 1% (or exceeded once within a specified return period, for example 100 years), resulting from one or more causes of loss affecting a portfolio of properties.

estimated based on the characteristics of the structure and the policy terms. More specifically, a probabilistic hurricane model contains the following four basic steps.

1. Assess the likelihood of events of various sizes, intensities and paths
2. Estimate the wind speeds at specific locations affected by each event
3. Estimate the damage to property, given the estimated wind speeds
4. Estimate the insured losses, given the damages.

A probabilistic hurricane model contains a comprehensive set of hypothetical events, each with an assigned probability. The event set is intended to provide a representative sampling of possible hurricane paths, sizes and intensities. Thus, it produces an estimate of the range of possible insured losses for any relevant location or geographical area. The statistical distribution of insured losses occurring at a particular location is reflective of the convolution of the four steps cited above.

At each of these steps, local validation is performed by comparing the model's predictions for a particular parameter to the available actual datasets. For example, the probability of an Atlantic hurricane making landfall in a particular coastal segment from the hypothetical sample can be compared to the actual number of landfalls since 1871, the beginning year of published records by NOAA.[2] Similarly, the model's probability of a hurricane with a particular size, path or intensity can be validated by comparison to historical hurricane records. The wind speed generated at a particular location for a simulated historical event can be compared to the actual observed wind speed. Finally, the predicted damages and insured losses to a particular type of structure subjected to a given wind speed can be compared to the actual damages and losses sustained at locations where that wind speed was present in a historical event.

At each step of the process, error is introduced to the extent that model results do not fully agree with actual observations. Model error is present because no model can precisely replicate an actual physical event. By definition, a model is a representation of the event; it seeks to capture the key underlying variables and their inter-relationships, leaving estimation errors from variables and inter-relationships not captured. Simulating a large number of hypothetical events can reduce certain of these errors. Some of the key contributors to hurricane model error are:

- In determining the likelihood of events of various sizes, intensities and paths
  - — limited availability of key parameters for a sufficient number of historical events
  - — limited availability of information on the historical frequency of rare events
  - — limited ability to predict changes in hurricane landfall frequency over time.

---

[2] The National Oceanic and Atmospheric Administration of the U S. Department of Commerce, publishes track and parameter information on hurricanes since 1871. In addition, there are numerous summaries and studies of prior documented storms. In recent years, there has also been research based on proxy approaches that derive past hurricane activity from geologic and biologic evidence.

- In estimating the wind speeds at specific locations affected by each event
  - limited availability of wind speed data for a sufficient number of locations for a sufficient number of historical events
  - limited ability to simulate the actual impact of land, vegetation and man-made objects on wind speeds
  - limited ability to simulate the possible variations in windfield shape (i.e., the distribution of wind velocity by distance and direction from the center), particularly including localized bursts of wind.

- In determining the damage to property
  - limited knowledge of precise types and values of property exposed at the time of the event
  - limited knowledge of the construction quality of those properties.

- In determining the insured losses
  - limited knowledge of claims adjusting practices of companies
  - limited availability of accurate historical insurance claims data in sufficient detail by location and coverage
  - limited knowledge of potential impact of governmental actions and demand surge
  - limitations in our ability to determine the portion of damage due to flood rather than wind.

These errors can be significant or modest in relation to the final results produced by the model. For example, Kelly and Zeng (Kelly and Zeng 1996) suggest that, based on their experience with one hurricane model, the errors introduced by the damage step are generally much less than a single order of magnitude while the errors introduced by the event steps can be several orders of magnitude. In other words, the model's estimate of expected losses for a particular risk might be off by 20% due to a mis-specified damage function, but those same expected losses might be off by 200% due to mis-estimation of the landfall probability.

Macro Validation Dataset

In the authors' view, public (and regulatory) acceptance of these models is hampered by the complexity of this layered validation approach, which leaves the outside user with an unclear picture of the overall goodness of fit between the model and historical data. The problem is only exacerbated when the model formulas and the validation results are treated as proprietary by the modelers. Accordingly, we set out to develop and publish a dataset permitting macro validation - one that would allow a lay person to compare the overall results of the model to an historical record. In addition to a comparison of model results to historical results, the dataset also demonstrates the limitations of the historical experience and data.

The macro validation dataset consists of the aggregate insured losses from each hurricane affecting the continental United States from 1900 through 1999. The dataset includes storms determined by NOAA to have caused hurricane conditions over land. Exhibit 1 lists these hurricanes[3] and shows their magnitude, as determined by NOAA, in each of the coastal states affected. The overall losses for each event have been allocated to county, based on estimates of relative loss within the state. The historical losses in each county have then been "trended" – adjusted from the conditions at the time to those existing today. Our work extends and improves upon similar work published by Pielke and Landsea (Pielke and Landsea 1998), which looks at total economic damages rather than insured losses and does not cover the entire 20th century.

Because the models are used primarily by the insurance industry, our focus was to estimate the aggregate insured losses directly sustained by the U.S. insurance industry. The same approaches described in the paper can be used to project total economic losses as well.

The remainder of this paper has two major sections. Section II describes the construction of the validation dataset, which consists of the losses from each historical event adjusted to 2000 cost and exposure levels. Section III illustrates the use of the dataset.

II. CONSTRUCTING THE VALIDATION DATA

Historical Losses

Data on the losses sustained from past hurricanes is available from a variety of public and private sources. The various data sources differ as to the types of costs included, the level of detail, and whether the figures are actual results or estimates.

The National Weather Service (NWS, which is part of NOAA) compiles data on the economic impact of each U.S. hurricane; that data is published annually in the *Monthly Weather Review*. A summary of this historical data from 1900 forward is presented in *Deadliest, Costliest, and Most Intense United States Hurricanes of This Century* (Hebert, Jarrell and Mayfield 1996). The data published by NWS are estimates based on surveys of the areas affected and consultations with experts, not a tabulation of actual costs incurred. The estimates include all direct costs stemming from the event, including insured losses, uninsured property losses, federal disaster assistance outlays, agriculture and environmental losses, etc. (Technically, the insured losses include some secondary costs due to the inclusion of business interruption and additional living expense claims.) Typically, the estimates for each event are not broken down by state or county. Separate estimates are made when a single hurricane makes more than one distinct landfall.

---

[3] The summary tables on Exhibit 1, Sheet 3, show total storms by category and state. Appendix A displays key statistics on hurricanes affecting Bermuda, Hawaii, Puerto Rico and USVI during the 20th century.

Property Claim Services, Inc. (PCS), a subsidiary of Insurance Services Office (ISO), prepares estimates of the direct insured losses for each natural catastrophe, including hurricanes. Their historical data extends back only to 1949. To be considered a catastrophe by PCS, the aggregate insured losses from the event must exceed a set dollar threshold. This threshold was originally set at $1 million; over time it has been raised to its current level of $25 million. The estimates published by PCS are based on surveys of insurers' reported loss activity, insurer market share data and a database of the number and types of structures by county. The current PCS practice is to prepare an initial loss estimate approximately two weeks after the event and to revise its estimates based on new information after subsequent 60 day periods until the estimate stabilizes, at which point no further revisions are made. Until the late 1980s, PCS estimates were rarely updated after 60 days and evidence suggests that these estimates often underestimated the total loss.

The PCS estimates are intended to include all insured losses paid directly by U.S. insurers under property and inland marine insurance coverages. This would include payment of the costs to repair or replace damaged property and contents, reimbursement for alternative housing while repairs are effected, and compensation for business interruption losses. The insurer's specific expenses for adjusting the claims are not included. The PCS estimates for each event are currently broken down by state, separately for personal property, commercial property and automobile, and also include the number of claims and the average payment.

Because they are prepared by different organizations using different source information, the NWS and PCS estimates of losses are not always consistent. Special studies have also been made in the past to collect actual insured losses for the industry. In a 1986 study, the All-Industry Research and Advisory Council (AIRAC) conducted a survey of insurers, asking them to provide their direct losses for the seven hurricanes occurring in 1983 and 1985 (AIRAC 1986). Responses were obtained from 95 insurers, who represented between 63% and 80% of the market share in the states affected. AIRAC then extrapolated the survey results to an industry-wide level based on the collective market shares of the respondents in the states affected by each event. (Collective market shares were based on premiums written by state.) In the AIRAC survey, insurers were requested to report their direct incurred losses including windstorm pool assessments, but excluding claim adjustment expenses. The AIRAC study indicated higher losses than the PCS estimates for 4 of the 7 hurricanes studied, including the 3 largest. In total for the seven storms, the AIRAC survey indicated losses of approximately $2.7 billion, 50% higher than the PCS estimate of $1.8 billion, as shown in Table 1.

TABLE 1

Comparison of PCS Estimates of Industry Losses to Estimates from the AIRAC Survey

Year	Hurricane	PCS Estimate	AIRAC Survey	Percent Difference
1983	Alicia	$675,000	$1,274,500	-47%
1985	Bob	13,000	9,946	31%
1985	Danny	37,000	24,509	51%
1985	Elena	543,000	622,050	-13%
1985	Gloria	418,000	618,299	-32%
1985	Juan	44,000	78,448	-44%
1985	Kate	77,000	67,830	14%
TOTAL		$1,807,000	$2,695,582	-49%

Certain state insurance departments also conduct studies of hurricane losses in their state. In the case of hurricane Andrew, the Florida Department of Insurance compiled the actual losses for the insurance industry. Under emergency rules promulgated by the Department, each insurer operating in the state was required to report their accumulated losses to the Department at the end of each quarter. The reported figures include only losses (i.e., not including costs of adjusting the claims), for Florida business only. Losses in Louisiana and elsewhere are not included.[4] The results as of March 31, 1994 were published in *The Journal of Reinsurance* (Lilly, Nicholson and Eastman 1994). In the aggregate, insurers reported 798,356 claims from hurricane Andrew, with a total dollar cost of approximately $16.1 billion. As of that date, insurers had paid out roughly 91.9% of that figure, with the balance representing their estimate of payments still to be made pending final adjustment. The final PCS estimate for Florida losses from Hurricane Andrew was $15 billion.

In constructing our validation dataset, we selected what we considered to be the best available estimate of the industry aggregate insured losses for each event. For events where no PCS or other direct estimate of insured losses was available, we estimated the insured losses as a percentage of the NWS/NOAA total loss estimate. There were 49 hurricanes for which no estimate of actual loss was available. This occurred only for weaker hurricanes that caused relatively small actual losses, generally those with under $1 million of actual losses prior to 1950. For these events, actual loss was estimated judgmentally. These judgmental estimates were selected to be consistent with estimates of total loss by year in Hebert, Jarrell and Mayfield

[4] Anecdotally, we would point out that insurance losses could be sustained by policyholders far away from the event. For example, in the case of hurricane Andrew an insurer sustained a loss by a Massachusetts policyholder who lost a camera while vacationing in Florida at the time. This loss would not be included in the figures quoted above.

(Hebert, Jarrell and Mayfield 1996). The normalized loss for these hurricanes represents only about 3% of the total normalized loss.

Allocation of Losses to County

Once a best estimate of the industry aggregate insured losses was selected, the losses were allocated to county. We devised a damage index for each county that reflected the estimated relative impact of the hurricane. The damage indices for all counties affected by an event were scaled such that, when multiplied by the number of housing units in the county at the time, the sum across all counties balanced to the selected industry aggregate insured loss.

The damage indices for an event are derived from the ToPCat hurricane model. The use of these indices means that the allocation of losses to county (and to state, prior to PCS estimates) is model-dependent. Nevertheless, the total insured loss estimates for each storm are not model dependent as they are balanced to the selected industry loss estimate.

Trending

The historical losses reflect the price levels and property exposure existing at the time of the event. If the same event were to happen today, the losses arising from that event would reflect

- today's price levels, reflecting the general inflation in price levels that occurred during the intervening period
- the current stock of properties and contents, reflecting the increase in the number of structures of various types, any increases in the average size or quality of the structures, and the greater amounts and value of the typical contents in the structures
- the current insurance system, including increases in the prevalence of insurance, the expansion of coverages, and changes in claim practices or the legal system governing how claims are settled.

Actuaries are accustomed to adjusting historical costs to current conditions by means of trend factors that account for changes in conditions during the intervening period. We developed trend factors to account for each of the three components above. Our goal was to adjust all historical losses forward to conditions prevailing in 2000.

The impact of monetary inflation was measured by reference to the Implicit Price Deflator (IPD) for Gross National Product, published by the Department of Commerce in their annual Economic Report to the President. An inflation trend factor was computed by dividing the estimated value of the IPD at year-end 2000 by the value at the time of the event. The IPD is only available back to 1950. For prior years, a 3.5% annual trend was assumed.

Of course, property values have increased by more than inflation. For example, the average size of houses and the amount of contents have gradually increased over time. The national growth in the value of property was measured using estimates of Fixed Reproducible Tangible Wealth (FRTW) published by the Department of Commerce's Bureau of Economic Analysis. FRTW measures the total value of all structures and equipment owned by businesses, institutions, and government as well as residential structures and durable goods owned by consumers. In this context, structures include buildings of all types, utilities, railroads, streets and highways, and military facilities. Similarly, equipment includes industrial machinery and office equipment, trucks, autos, ships, and boats. While FRTW includes some elements not entirely relevant to property insurance such as military facilities and highways, these elements represented less than 10% of the total as of year-end 1995.

FRTW estimates are prepared annually; time-series data is presented on several different bases. We utilized the Real Net Stock of FRTW series, which is net of depreciation, and adjusted to 1992 dollar levels such that it accounts only for real and not inflationary growth in the net value of property over time. A national property growth factor was computed by dividing the estimated value of the Real Net Stock of FRTW at year-end 2000 by the corresponding value at the time of the event. This growth factor accounts for aggregate growth in property values due to population growth and increases in per capita wealth. The selected FRTW series is only available back to 1925. For prior years, we assumed a 2.5% per year trend.

The national growth in property exposure has been far from uniform geographically. The general migration of the U.S. population towards the South and West over the last several decades has been well publicized. Of particular relevance to potential hurricane losses is the increased concentration of people and property in vulnerable coastal locations.

Pielke and Landsea (Pielke and Landsea 1998) have suggested that the national property growth factor be adjusted based on relative growth of the population in the affected region versus the nation as a whole. They introduce a population adjustment equal to the ratio of the growth in population in the affected coastal counties to the growth in population nationally. While this approach reasonably captures the migration of the U.S. population to the Sunbelt, it fails to take into account the explosive growth in vacation homes. (Census population data accounts for people at the location of their principal residence.) This issue is particularly significant because a large number of vacation homes are located in coastal resort areas: Cape Cod, Long Island, Cape Hatterras, Florida, etc.

We improve upon Pielke and Landsea's approach by using the growth in the total number of housing units in each county during the time period for which it is available, rather than the growth in population. Housing unit data is available from the Census, back to 1940. (County data from the decennial census was interpolated to obtain annual housing unit estimates for each county. Prior to 1940, we used population statistics to estimate housing units.)

A second improvement relates to the way in which the county data is used. Pielke and Landsea (Pielke and Landsea 1998) identified the coastal counties that were affected by each event and based their geographic adjustment on the aggregate change in population for all counties combined. Because we estimated the insured loss by county, we were able to weight the growth by relative damage in each county.

Since we are adjusting insured losses, a final adjustment was necessary to account for changes in the insurance system. Ideally, this adjustment should account for each of the following.

- *Changes in the prevalence of insurance coverage.* Coverage for the wind peril is fairly universal today, primarily because mortgage lenders require it. (This requirement does not exist for earthquake insurance, resulting in significantly lower market penetration for that coverage, even in earthquake-prone areas.) Property that is uninsured tends to be lower valued. Prior to the introduction of multiple peril policies in the 1960s, wind coverage was far less universal. The introduction of FAIR plans and wind pools has also contributed to more universal coverage.

- *Changes in the level and structure of coverage.* Competition has led to gradual increases in the level of coverage offered by standard insurance policies. For example, coverage for contents, generally written as a standard percentage of building coverage on personal lines policies, has increased over time. More significantly, there has also been a longer-term trend away from actual cash value to replacement cost coverage. This shift has been widespread in homeowners; even some business-owners is now written on a replacement cost basis. Conning (Conning & Company 1996) has pointed out that this change in coverage significantly increases the insurer's exposure, essentially changing it from a net (of depreciation) to a gross value basis. One coverage trend has acted to reduce insurers hurricane exposure in recent years. Subsequent to Hurricane Andrew, there was a significant increase in required deductibles in coastal areas. While individuals have tended to resist voluntary increases in retentions, there has been a longer-term trend toward larger self-insured retentions in the commercial insurance sector.

- *Changes in the typical practices regarding claim settlements.* While this element may be the hardest to specify, industry professionals believe that policyholders have a greater propensity to file claims, particularly claims relating to minor or consequential damage. At the same time, insurers are more willing to interpret the coverage in a manner favorable to the insured (contrary to public perception), in the interests of customer satisfaction, particularly after a catastrophe.

Taken collectively, all of these factors work to increase the extent of economic losses covered by insurance, particularly as one goes further back in time. The insurance utilization index was derived from a review of ratios of PCS insured loss estimates to NOAA economic loss estimates from 1949 through 1995. The data and selected insurance utilization index are compared in the graph on Appendix B, Exhibit 2. The selected index from 1950 through 1995 was based on a linear least squares fit of the data. The fit produced a line from approximately 21% in 1950 through 55% in 1995. From 1995 through 2000, the insurance utilization rate was kept at a constant 55% to judgmentally reflect the increasing use of deductibles. Prior to 1950, a linear trend from 10% in 1900 through 21% in 1950 was judgmentally selected. As total economic losses were used as the starting point for normalization prior to 1949, this latter assumption has virtually no impact on normalized losses.

Appendix B, Exhibit 1 displays the historical growth rates in the IPD and FRTW indexes as well as the national growth in population and housing units.

Mathematically, the trend procedure can be expressed as follows:

$$L_{c,2000} = L_{c,y} \text{ x } \left(\frac{IPD_{2000}}{IPD_y}\right) \text{x} \left(\frac{FRTW_{2000}}{FRTW_y}\right) \text{x} \left(\frac{HU_{c,2000}/HU_{c,y}}{HU_{2000}/HU_y}\right) \text{x} \left(\frac{INS_{2000}}{INS_y}\right)$$

Where:

$L_{c,y}$	is the insured loss in county c from an event in year y
IPD_y	is the value of the Implicit Price Deflator for year y
$FRTW_y$	is the Real Net Stock of Fixed Reproducible Tangible Wealth for year y
$HU_{c,y}$	is the estimated number of Census Housing Units in county c in year y
INS_y	in the insurance utilization index for year y

Limitations of the dataset

We believe that the validation dataset produced by the normalization process described above is useful for comparing the results of U.S. hurricane models to the historical record. The dataset provides a macro tool that can be used by model users with limited knowledge of the detailed assumptions underlying the model. Nevertheless, it should be expected that probabilistic model results will vary from the results of the normalization process. The causes of this variation can be segregated into two types: variations caused by limitations in the normalization model, and variations caused by basic differences between a historical normalization process and a probabilistic model. A summary of the causes of each type is outlined below.

- Limitations of the normalization process itself (these limitations would also relate to comparisons of normalized and modeled historical storm results)
  - — unavailability of insured loss estimates prior to the inception of PCS estimates in 1949
  - — inaccuracies in the historical PCS insured loss estimates (as previously noted, the AIRAC study in 1986 and the Florida Department of Insurance study of Hurricane Andrew in 1992 both indicated significantly different levels of industry losses than the PCS estimates)

- leveraging in the trending procedure (small changes in the initial estimate of the insured loss or its allocation to county can produce large changes in the normalized amount for events that occurred many years ago; this distortion should be less significant at the statewide level or for groups of neighboring counties)
- trending of exposures based solely on housing units (normalized losses in counties with commercial property growth significantly different than housing unit growth will be distorted)

- Basic differences between historical normalization and probabilistic models
  - probabilistic models provide a representative sampling of possible hurricane paths, sizes and intensities, which can produce results that differ significantly from the results of one hundred-year period that are influenced greatly by the location of the 5 or 10 largest or most intense storms
  - probabilistic model industry loss estimates are dependent on the accuracy of the modeler's estimate of total insured property exposures by ZIP code or county that are used in the modeling to estimate industry loss (these industry exposure sets are independently developed by modelers, or may be developed by users, based on insurance industry or external statistics on property values)
  - probabilistic models may include tropical storms that do not reach hurricane strength or strafing hurricanes that do not produce hurricane winds over land (these differences can distort loss comparisons as well as frequency comparisons)

Results

Exhibit 2 presents an illustrative calculation of losses in Mississippi from Hurricane Camille. The inputs are the year of the event, the estimated total losses for the event, by state (from PCS) and the damage index for each county. To illustrate how inflation, real growth in property values, population migration, insurance utilization and housing units combine to increase the level of economic losses from a hurricane, we will look at the figures for the two counties contributing most to the Mississippi losses: Hancock and Harrison. Since 1969, housing units have grown by 222.8% in Hancock and 90.7% in Harrison. The normalization process brings the Hancock losses up by 2716%, from approximately $20 million to $549 million, while the Harrison losses increase by 1604%. The Hancock increase is attributed to:

Inflation	297.4%
Growth in wealth per capita (2.317 ÷ 1.703)	36.1%
Growth in insurance utilization	55.6%
Growth in housing units	222.8%

Thus, in Hancock County, the impact of inflation (297.4%) is less than the combined impact of the other three factors (584% = (1.361 x 1.556 x 3.228)-1), the most important of which is the growth in the number of housing units.

Exhibit 3 summarizes the estimated actual and normalized losses for hurricanes affecting the U.S. during the 20th century. The normalized losses for these 164 hurricanes average $1.75 billion per storm, or $2.87 billion per year. The resulting size of loss distribution by Saffir-Simpson category on Exhibit 3, Sheet 4 shows the impact of storm severity on insurance losses. While only about 9% of historical events were category 4 hurricanes, those events produced 55% of the normalized losses. Interestingly, the category 5 hurricanes have not produced a similarly skewed impact because the only two such events (#2 in 1935 and Camille in 1969) did not hit densely populated areas.

Exhibit 3, Sheet 4 also shows the variation in normalized loss by decade, most notably the high losses in the twenties and the relatively low losses in the seventies and eighties.

III USING THE VALIDATION DATA

Severity Distributions by State

Exhibit 4 displays annual aggregate (Sheet 1) and maximum single occurrence (Sheet 2) distributions by state based on the normalized losses from 1900 to 1999. Due to the low probability of having more than one hurricane per year in most states, the results in Sheets 1 and 2 are quite similar. Florida, with almost 50% of the expected annual losses, and Texas, with over 21%, dominate the results. The total annual aggregate distributions at the longer return periods (20 years and greater) are also driven by the worst storms in those two states.

As 100 years is not a sufficiently long time period to credibly determine the likely loss levels at the longer return periods, random elements are evident in the state distributions. For example, the 100-year loss for South Carolina, Hurricane Hugo in 1989, is approximately 10 times the 100-year loss in Georgia, Hurricane Opal in1995. Georgia was not hit heavily in the 20thcentury, having had no landfalling events, but saw several major hurricanes in the 19th century. On a probabilistic basis, it is reasonable to expect the 100-year loss in Georgia to be somewhat closer to the South Carolina 100-year loss.

The normalized results by state are compared to those of a hypothetical representative probabilistic hurricane model ("Model T") in Exhibit 6, Sheets 1 and 2. Sheet 1 compares normalized and modeled frequency and severity distributions by Saffir-Simpson category and by return period for Texas, Florida and countrywide. Sheet 2 compares normalized and modeled expected losses by state. Based on the Model T indications, Georgia, New Jersey and New York were relatively lucky during the 20th century, while Texas was the most unlucky. Comparisons such as those in Exhibit 6 could be used to learn more about the assumptions behind a probabilistic model. For example, in this case it would be useful to learn the answers to questions such as:

- What data are the Model T frequency distributions based on, and why do they differ from the 20th century distributions?
- What are the paths and Saffir-Simpson categories of the typical 50 year and 100 year return events in Model T, compared to the worst events by state during the 20th century?
- Why are the Model T expected losses in Texas so much lower and New York and New Jersey so much higher than the normalized 20th century expected losses?
- How do these and other key differences from the 20th century storm set affect the results of Model T on a specific insurer's portfolio?

Severity Distributions by County

Exhibit 5 displays annual aggregate loss distributions for counties with significant annual expected losses in Texas and Florida. Random elements are even more evident at the county level. For example, Dade County has expected losses over 3 times expected losses in Broward County and over 5 times those in Palm Beach County, Florida, due to the influence of Hurricane Andrew and storm number 6 of 1926.

These results could be compared to the results of a probabilistic model to determine how the model's expected losses vary from historical results in these counties. For example, Model T indicates expected losses in Dade County 27% higher than in Broward County and 36% higher than in Palm Beach County. Of course, as one looks at smaller geographic areas (e.g., county rather than state), one would expect larger differences between a model and the historical results of one hundred-year period.

Estimates of Losses from Historical Events

Exhibit 6, Sheet 3 compares the normalized losses from the 50 largest events of the 20th century to the Model T results for those same events. Here we see evidence that modeled individual storm estimates often differ significantly from the normalized amounts. Differences of over 50% occur on 18 of the 50 storms. These differences occur primarily on storms prior to the advent of PCS estimates in 1949. Only 2 of the 18 (Hurricane King in 1950 and Hurricane Donna in 1960) have normalized estimates based on PCS. These differences indicate the uncertainty in both normalizing and modeling these older storms.

In conclusion, the normalized hurricane loss database provides a variety of tools for hurricane model users to perform macro validation tests of model assumptions. In keeping with the spirit of this call for papers on data, the authors will provide interested readers with an electronic copy of the normalized loss database by event and county. We trust that future research will expand the scope of hurricane loss data to include not only hurricanes of the 21st century, but improvements to this 20th century database, and perhaps also the addition of estimates of hurricane losses in prior centuries.

IV. REFERENCES

All-Industry Research Advisory Council. 1986. *Catastrophe Losses, How the Insurance System Would Handle Two $7 Billion Hurricanes.*

Conning & Company. 1996. *Homeowners Insurance: The Problem is the Product.* Hartford, CT.

Hebert, P.J., J.D. Jarrell, and M. Mayfield. Updated February 1996. *Deadliest, Costliest, and Most Intense United States Hurricanes of This Century (and other Frequently Requested Hurricane Facts).* NOAA Technical Memorandum NWS TPC-1.

Internet Files: NHC.NOAA.gov/; NCEP.NOAA.gov/.

Kelly, P.J., and L. Zeng. 1996. *The Engineering, Statistical, and Scientific Validity of EQECAT USWIND Modeling Software.* Page.2. Presented at the ACI Conference for Catastrophe Reinsurance. New York, NY.

Lilly III, C.C., J.E. Nicholson, and K. Eastman. 1994. Hurricane Andrew: Insurer Losses and Concentration. *Journal of Reinsurance* Volume 1, Number 3, p. 34.

Neumann, Charles J., Brian R. Jarvinen, Colin J. McAdie and Joe D. Elms. 1993. *Tropical Cyclones of the North Atlantic Ocean, 1871–1992.* 4th revision. Asheville, NC: National Climatic Data Center.

Pielke, Jr., R.A., and C.W. Landsea. 1998. Normalized Hurricane Damages in the United States: 1925-1995. *Weather and Forecasting* Number 13, p. 621.

Tucker, Terry. Amended1995. *Beware the Hurricane.* Bermuda: The Island Press Limited.

Hurricanes Affecting the Continental U.S. 1900 - 1999

	Hurricane	Date of	Category and Coastal States Affected																									
Year	Number/ Name	First US Landfall	TX So	TX Ce	TX No	TX	LA	MS	AL	FL NW	FL SW	FL SE	FL NE	FL	GA	SC	NC	VA	MD	DE	NJ	NY	CT	RI	MA	NH	ME	CW
1900	1	08-Sep			4	4																						4
1901	3	10-Jul															1											1
1901	4	14-Aug					2	2																				2
1903	3	11-Sep								1		2		2														2
1903	4	16-Sep																			1	1	1					1
1904	2	14-Sep														1	L											1
1906	2	16-Jun										1		1														1
1906	4	17-Sep														3	3											3
1906	5	27-Sep						3	3	L																		3
1906	8	17-Oct									L	2		2														2
1908	2	30-Jul															1											1
1909	3	21-Jul			3	3																						3
1909	5	27-Aug	2			2																						2
1909	7	20-Sep					4																					4
1909	9	11-Oct									L	3		3														3
1910	2	14-Sep	2			2																						2
1910	4	17-Oct									3	L		3														3
1911	1	11-Aug							1	1				1														1
1911	2	27-Aug													2	2												2
1912	3	13-Sep							1																			1
1912	5	16-Oct	1			1																						1
1913	1	27-Jun	1			1																						1
1913	2	02-Sep															1											1
1915	2	17-Aug			4	4																						4
1915	4	04-Sep								1				1														1
1915	5	29-Sep					4																					4
1916	1	05-Jul					L	3	3																			3
1916	2	21-Jul																							1			1
1916	3	14-Jul														1												1
1916	4	18-Aug	3			3																						3
1916	13	18-Oct							2	2				2														2
1916	14	15-Nov									1			1														1
1917	3	28-Sep								3				3														3
1918	1	06-Aug			L		3																					3
1919	2	14-Sep	4			4					4			4														4
1920	2	21-Sep					2																					2
1920	3	22-Sep															1											1
1921	1	22-Jun		2		2																						2
1921	6	25-Oct								L	3		2	3														3
1923	3	15-Oct					1																					1
1924	4	15-Sep								1				1														1
1924	7	20-Oct									1	L		1														1
1925	2	01-Dec									1		L	1														1
1926	1	27-Jul											2	2														2
1926	3	25-Aug					3																					3
1926	6	18-Sep						L	3	3	3	4		4														4
1928	1	07-Aug										2	L	2														2
1928	4	16-Sep										4	2	4	1	1												4
1929	1	28-Jun		1		1																						1
1929	2	28-Sep								2		3		3														3
1932	2	13-Aug			4	4																						4
1932	3	01-Sep							1																			1
1933	5	30-Jul	2			2						1		1														2
1933	8	23-Aug															2	2										2
1933	11	04-Sep	3			3																						3
1933	12	03-Sep										3		3														3
1933	13	16-Sep															3											3
1934	2	16-Jun					3																					3
1934	3	25-Jul	2			2																						2
1935	2	03-Sep								2	5			5														5
1935	6	04-Nov									L	2		2														2
1936	3	27-Jun	1	L		1																						1
1936	5	31-Jul								3				3														3
1936	13	18-Sep															2											2
1938	2	14-Aug					1																					1
1938	4	21-Sep																				3	3	3	3			3
1939	2	11-Aug								1		1		1														1
1940	2	07-Aug			2	2	2																					2
1940	3	11-Aug													2	2												2
1941	2	23-Sep			3	3																						3

Hurricanes Affecting the Continental U.S. 1900 - 1999

	Hurricane	Date of	Category and Coastal States Affected																									
Year	Number/ Name	First US Landfall	TX So	TX Ce	TX No	TX	LA	MS	AL	FL NW	FL SW	FL SE	FL NE	FL	GA	SC	NC	VA	MD	DE	NJ	NY	CT	RI	MA	NH	ME	CW
1941	5	06-Oct								2	2	2		2														2
1942	1	21-Aug			1	1																						1
1942	2	29-Aug		3	L	3																						3
1943	1	26-Jul			2	2																						2
1944	3	01-Aug															1											1
1944	7	14-Sep															3	3				3	3	3	2			3
1944	11	18-Oct									3		2	3														3
1945	1	24-Jun								1				1														1
1945	5	26-Aug		2	L	2																						2
1945	9	15-Sep										3		3														3
1946	5	07-Oct									1			1														1
1947	3	24-Aug			1	1																						1
1947	4	17-Sep					3	3			2	4		4														4
1947	8	11-Oct									1	1		1	2	2												2
1948	5	03-Sep					1																					1
1948	7	21-Sep									3	2		3														3
1948	8	05-Oct									L	2		2														2
1949	1	24-Aug															1											1
1949	2	26-Aug										3	L	3	L													3
1949	10	03-Oct			2	2																						2
1950	Baker	30-Aug							1																			1
1950	Easy	04-Sep								3				3														3
1950	King	17-Oct										3	L	3														3
1952	Able	30-Aug														1												1
1953	Barbara	13-Aug															1											1
1953	Carol	07-Sep																									1	1
1953	Florence	26-Sep								1				1														1
1954	Carol	31-Aug															2					3	3	3	L	L	L	3
1954	Edna	11-Sep																					L	L	3	L	1	3
1954	Hazel	15-Oct														4	4	L	2	L	L	L						4
1955	Connie	12-Aug															3	1	L									3
1955	Diane	17-Aug															1											1
1955	Ione	19-Sep															3											3
1956	Flossy	24-Sep					2			1				1														2
1957	Audrey	27-Jun			4	4	4																					4
1959	Cindy	08-Jul														1												1
1959	Debra	24-Jul			1	1																						1
1959	Gracie	29-Sep														3												3
1960	Donna	09-Sep									4		2	4			3					3	2	2	1	1	1	4
1960	Ethel	15-Sep						1																				1
1961	Carla	11-Sep		4	L	4																						4
1963	Cindy	17-Sep			1	1																						1
1964	Cleo	26-Aug										2	L	2														2
1964	Dora	09-Sep											2	2	L													2
1964	Hilda	03-Oct					3																					3
1964	Isbell	14-Oct									2	2		2														2
1965	Betsy	08-Sep					3	L				3		3														3
1966	Alma	09-Jun								2				2														2
1966	Inez	04-Oct									1	L		1														1
1967	Beulah	20-Sep	3			3																						3
1968	Gladys	18-Oct								2			1	2														2
1969	Camille	17-Aug					5	5																				5
1969	Gerda	09-Sep																									1	1
1970	Celia	03-Aug	3	L		3																						3
1971	Edith	16-Sep					2																					2
1971	Fern	09-Sep		1		1																						1
1971	Ginger	30-Sep															1											1
1972	Agnes	19-Jun								1				1				L	L			1	1					1
1974	Carmen	07-Sep					3																					3
1975	Eloise	23-Sep							L	3				3														3
1976	Belle	09-Aug																				1						1
1977	Babe	04-Sep					1																					1
1979	Bob	11-Jul					1																					1
1979	David	03-Sep										2	2	2	2	2	L											2
1979	Frederic	12-Sep						3	3	L																		3
1980	Allen	09-Aug	3			3																						3
1983	Alicia	17-Aug			3	3																						3
1984	Diana	11-Sep															3											3
1985	Bob	24-Jul														1												1
1985	Danny	15-Aug					1																					1

Hurricanes Affecting the Continental U.S. 1900 - 1999

Year	Hurricane Number/ Name	Date of First US Landfall	TX So	TX Ce	TX No	TX	LA	MS	AL	FL NW	FL SW	FL SE	FL NE	FL	GA	SC	NC	VA	MD	DE	NJ	NY	CT	RI	MA	NH	ME	CW
			Category and Coastal States Affected																									
1985	Elena	01-Sep						3	3	3				3														3
1985	Gloria	27-Sep															3				L	3	2	L	L	2	1	3
1985	Juan	28-Oct					1																					1
1985	Kate	21-Nov								2				2	L													2
1986	Bonnie	26-Jun			1	1																						1
1986	Charley	17-Aug															1	1										1
1987	Floyd	12-Oct									1			1														1
1988	Florence	09-Sep					1																					1
1989	Chantal	01-Aug			1	1																						1
1989	Hugo	21-Sep														4	L											4
1989	Jerry	15-Oct			1	1																						1
1991	Bob	19-Aug																				2	2	2	2		L	2
1992	Andrew	24-Aug					3				3	4		4														4
1993	Emily	01-Sep															3											3
1995	Erin	01-Aug							L	L		1	L	1														1
1995	Opal	04-Oct							L	3				3	L													3
1996	Bertha	12-Jul															2	L										2
1996	Fran	05-Sep														L	3	L	L									3
1997	Danny	18-Jul					1		L																			1
1998	Bonnie	26-Aug														L	2	1										2
1998	Earl	02-Sep								1				1														1
1998	Georges	28-Sep					L	2	L		2			2														2
1999	Bret	22-Aug	3			3																						3
1999	Floyd	16-Sep											L			L	2	L	L	L	L	L	L					2
	Number of Hurricanes Affecting, by Category:																											
	1		3	2	7	12	9	1	4	10	7	5	1	19	1	6	10	3			1	3	2		2	1	5	61
	2		4	2	3	9	5	2	1	7	4	10	7	16	4	4	6	1	1			1	3	2	2	1		38
	3		6	1	3	10	8	5	5	7	6	7		17		2	10	1				5	3	3	2			48
	4		1	1	4	6	3				2	4		6		2	1											15
	5						1	1			1			1														2
	Total		14	6	17	37	26	9	10	24	20	26	8	59	5	14	27	5	1	0	1	9	8	5	6	2	5	164
	Additional areas with normalized damage greater than $25 million:																											
	L		0	2	4	1	2	2	5	4	4	3	7	3	4	3	3	5	4	2	3	2	2	2	2	2	2	0
	Number of Hurricanes making First Landfall, by Category:																											
	1		3	2	7	12	9	1	4	6	7	4		17		5	10				1	1			1		2	63
	2		3	2	3	8	4	1		4	1	7	3	15		2	5							1				36
	3		6	1	3	10	5	3	1	5	4	7		16		2	6					3		1	1			48
	4		1	1	4	6	2				1	4		5		1	1											15
	5							1			1			1														2
	Total		13	6	17	36	20	6	5	15	14	22	3	54	0	10	22	0	0	0	1	4	0	2	2	0	2	164

Notes:

Coastal states affected, and category designations according to Saffir-Simpson Hurricane Scale, based on Neumann (Neumann, Jarvinen, McAdie and Elms, 1993) through 1992, and on NOAA summary reports for 1993-1999. States "affected" reflects NOAA's judgment as to which areas received hurricane conditions at the intensity of the defined Saffir-Simpson category. In some cases, the conditions may have existed only in very localized areas and may not have existed in areas that contained significant amounts of insured property. Additional states with normalized losses greater than $25 million noted by 'L'
First landfall indicated by italics (strafing of coastal islands not considered as first landfall if subsequent landfall more significant)

Saffir-Simpson Scale Number (Category)	Central Pressure (Millibars)	OR Winds (MPH)	OR Surge (Feet)
1	>979	74-95	4-5
2	965 - 979	96-110	6-8
3	945 - 964	111-130	9-12
4	920 - 944	131-155	13-18
5	<920	>155	>18

Coastal County Definitions:

Texas South is Cameron to Nueces Counties
Texas Central is San Patricia to Matagorda Counties
Texas North is Brazoria to Orange Counties

Florida Northwest is Escambia to Pasco Counties
Florida Southwest is Pinellas to Monroe Counties
Florida Southeast is Dade to Indian River Counties
Florida Northeast is Brevard to Nassau Counties

Normalization of Catastrophe Losses for Inflation and Real Growth in Property
Hurricane Camille - August 17, 1969

State	County	Housing Units At Time Of Event 1969	Damage Index	Estimated Losses (000's) At Time Of Event 1969	Growth in Number of Housing Units	Overall Adjustment Factor	Estimated Losses (000's) Adjusted to 2000
MS	Amite County	4,353	0.6%	$ 26	38.4%	1164%	$ 306
MS	Attala County	6,586	1.0%	69	19.2%	1003%	690
MS	Carroll County	3,017	1.1%	34	53.7%	1293%	434
MS	Choctaw County	2,824	0.1%	4	33.8%	1126%	42
MS	Clarke County	5,077	0.4%	21	50.5%	1266%	268
MS	Copiah County	7,652	1.0%	77	45.9%	1227%	947
MS	Covington County	4,207	15.9%	668	74.6%	1469%	9,811
MS	Forrest County	18,642	14.0%	2,601	71.0%	1439%	37,417
MS	George County	3,860	7.2%	279	113.1%	1792%	5,002
MS	Greene County	2,691	2.8%	76	87.8%	1580%	1,205
MS	Grenada County	6,412	0.8%	53	49.4%	1257%	669
MS	Hancock County	7,230	279.3%	20,198	222.8%	2716%	548,553
MS	Harrison County	40,778	206.9%	84,387	90.7%	1604%	1,353,784
MS	Hinds County	65,870	1.7%	1,113	51.6%	1275%	14,190
MS	Holmes County	7,145	2.2%	157	12.9%	949%	1,495
MS	Humphreys County	4,314	0.1%	5	-8.3%	771%	38
MS	Jackson County	26,463	37.2%	9,856	111.6%	1780%	175,443
MS	Jasper County	4,956	1.5%	74	42.9%	1202%	889
MS	Jefferson Davis County	3,865	21.9%	845	40.1%	1178%	9,959
MS	Jones County	18,104	3.5%	635	47.5%	1241%	7,880
MS	Lamar County	4,842	28.1%	1,362	215.8%	2656%	36,172
MS	Lawrence County	3,530	7.1%	252	58.6%	1334%	3,358
MS	Leake County	5,742	1.2%	68	48.3%	1248%	842
MS	Leflore County	13,048	0.7%	95	6.5%	896%	853
MS	Lincoln County	8,591	0.7%	59	54.9%	1303%	771
MS	Madison County	8,202	3.8%	311	289.4%	3276%	10,175
MS	Marion County	7,305	28.9%	2,108	47.6%	1242%	26,168
MS	Montgomery County	4,210	0.8%	35	22.1%	1027%	355
MS	Neshoba County	6,991	0.1%	10	66.0%	1396%	143
MS	Newton County	6,493	0.6%	40	40.6%	1183%	469
MS	Panola County	7,932	0.2%	19	71.1%	1439%	276
MS	Pearl River County	8,637	101.3%	8,753	136.2%	1987%	173,896
MS	Perry County	2,819	8.2%	232	81.5%	1527%	3,543
MS	Pike County	10,625	0.7%	75	53.0%	1287%	964
MS	Rankin County	11,753	7.3%	856	265.6%	3075%	26,319
MS	Scott County	6,581	3.9%	257	59.1%	1338%	3,437
MS	Simpson County	6,378	13.8%	882	64.6%	1384%	12,206
MS	Smith County	4,427	7.3%	321	39.9%	1177%	3,781
MS	Stone County	2,450	28.2%	690	129.4%	1930%	13,324
MS	Tallahatchie County	6,241	0.5%	31	-11.4%	745%	231
MS	Walthall County	4,006	6.3%	253	45.7%	1226%	3,096
MS	Wayne County	5,033	0.9%	44	64.2%	1381%	606
MS	Webster County	3,378	0.3%	9	36.2%	1146%	102
MS	Winston County	5,836	0.1%	5	34.1%	1128%	54
MS	Yalobusha County	4,130	0.4%	18	38.0%	1161%	204
MS	Yazoo County	8,700	0.5%	39	11.2%	935%	367
	Mississippi Total			138,000	114.6%	1805%	2,490,730
	Alabama			2,000	101.8%	1698%	33,950
	Florida			1,000	173.1%	2297%	22,972
	Louisiana			25,000	91.2%	1609%	402,137
	Event Total			166,000			2,949,789

Countrywide:	Change in Price Level - GNP Deflator	297.4%
	Real Growth in National Wealth	131.7%
	Growth in Insurance Utilization	55.6%
	Growth in Number of Housing Units	70.3%

Hurricane Loss Estimates
Continental U.S. 1900 - 1999
Dollars in Thousands

Year	Hurricane Number/ Name	Total Estimated Actual Loss at Time of Event: Economic	Insurance Utilization	Insured	Source	Insured Loss Normalized To 2000	Max Loss State/Region	Max Category
1900	1	$ 30,000	10.0%	$ 1,500	NOAA	$ 16,485,683	TX - No	4
1901	3	100	10.2%	10	NOAA	76,846	NC	1
1901	4	925	10.2%	95	NOAA	366,142	LA	2
1903	3	800	10.7%	85	NOAA	2,124,106	FL - SE	2
1903	4	200	10.7%	21	NOAA	61,970	NJ	1
1904	2	2,000	10.9%	218	NOAA	646,193	SC	1
1906	2	100	11.3%	11	NOAA	894,836	FL - SE	1
1906	4	1,500	11.3%	170	NOAA	525,681	NC	3
1906	5	1,500	11.3%	170	NOAA	662,658	AL	3
1906	8	100	11.3%	11	NOAA	687,544	FL - SE	2
1908	2	100	11.8%	12	NOAA	37,659	NC	1
1909	3	1,900	12.0%	228	NOAA	1,119,560	TX - No	3
1909	5	100	12.0%	12	NOAA	87,098	TX - So	2
1909	7	1,100	12.0%	132	NOAA	189,900	LA	4
1909	9	5,000	12.0%	599	NOAA	7,976,601	FL - SE	3
1910	2	100	12.2%	12	NOAA	75,760	TX - So	2
1910	4	1,000	12.2%	122	NOAA	2,735,157	FL - SW	3
1911	1	675	12.4%	84	NOAA	438,296	FL - NW	1
1911	2	325	12.4%	40	NOAA	58,145	SC	2
1912	3	100	12.6%	13	NOAA	27,091	AL	1
1912	5	100	12.6%	13	NOAA	65,024	TX - So	1
1913	1	100	12.9%	13	NOAA	66,228	TX - So	1
1913	2	3,000	12.9%	386	NOAA	534,237	NC	1
1915	2	50,000	13.3%	4,988	NOAA	16,146,375	TX - No	4
1915	4	100	13.3%	13	NOAA	43,577	FL - NW	1
1915	5	13,000	13.3%	1,729	NOAA	1,709,809	LA	4
1916	1	30,000	13.5%	2,028	NOAA	3,096,434	MS	3
1916	2	125	13.5%	17	NOAA	15,474	MA	1
1916	3	100	13.5%	14	NOAA	17,866	SC	1
1916	4	350	13.5%	47	NOAA	147,702	TX - So	3
1916	13	1,125	13.5%	152	NOAA	208,433	FL - NW	2
1916	14	300	13.5%	41	NOAA	65,139	FL - SW	1
1917	3	100	13.7%	14	NOAA	28,690	FL - NW	3
1918	1	5,000	14.0%	698	NOAA	775,971	LA	3
1919	2	22,000	14.2%	3,120	NOAA	10,009,409	FL - SW	4
1920	2	3,000	14.4%	432	NOAA	348,405	LA	2
1920	3	100	14.4%	14	NOAA	18,497	NC	1
1921	1	275	14.6%	40	NOAA	31,069	TX - No, Ce	2
1921	6	2,725	14.6%	398	NOAA	1,624,995	FL - SW	3
1923	3	100	15.1%	15	NOAA	9,557	LA	1
1924	4	100	15.3%	15	NOAA	12,256	FL - NW	1
1924	7	100	15.3%	15	NOAA	86,278	FL - SE	1
1925	2	250	15.5%	39	NOAA	155,351	FL - SW	1
1926	1	3,000	15.7%	472	NOAA	1,755,434	FL - NE	2
1926	3	4,000	15.7%	629	NOAA	305,313	LA	3
1926	6	105,000	15.7%	16,506	NOAA	49,728,840	FL - SE	4
1928	1	250	16.2%	40	NOAA	132,787	FL - SE	2
1928	4	25,000	16.2%	4,040	NOAA	9,816,472	FL - SE	4
1929	1	250	16.4%	41	NOAA	18,946	TX - Ce	1
1929	2	975	16.4%	160	NOAA	356,558	FL - SE	3
1932	2	7,500	17.0%	1,278	NOAA	836,911	TX - No	4
1932	3	250	17.0%	43	NOAA	32,860	AL	1

Exhibit 3
Sheet 2

Hurricane Loss Estimates
Continental U.S. 1900 - 1999
Dollars in Thousands

Year	Hurricane Number/ Name	Total Estimated Actual Loss at Time of Event: Economic	Insurance Utilization	Insured	Source	Insured Loss Normalized To 2000	Max Loss State/Region	Max Category
1933	5	$ 250	17.3%	$ 43	NOAA	$ 67,732	FL - NE	1
1933	8	17,000	17.3%	2,934	NOAA	1,356,989	VA	2
1933	11	1,000	17.3%	173	NOAA	368,245	TX - So	2
1933	12	12,000	17.3%	2,071	NOAA	1,163,819	FL -SE	3
1933	13	1,000	17.3%	173	NOAA	75,739	NC	3
1934	2	2,600	17.5%	454	NOAA	133,959	LA	3
1934	3	250	17.5%	44	NOAA	17,976	TX -So	2
1935	2	6,000	17.7%	1,062	NOAA	1,191,386	FL - SW	5
1935	6	5,500	17.7%	974	NOAA	1,371,030	FL - SE	2
1936	3	250	17.9%	45	NOAA	17,658	TX - So	1
1936	5	250	17.9%	45	NOAA	20,289	FL - NW	3
1936	13	250	17.9%	45	NOAA	18,891	VA	2
1938	2	250	18.4%	46	NOAA	9,005	LA	1
1938	4	306,000	18.4%	56,182	NOAA	9,965,606	CT	3
1939	2	250	18.6%	46	NOAA	41,746	FL - NE	1
1940	2	250	18.8%	47	NOAA	8,223	TX - No	2
1940	3	7,000	18.8%	1,316	NOAA	293,910	SC	2
1941	2	950	19.0%	181	NOAA	64,533	TX - No	3
1941	5	7,050	19.0%	1,341	NOAA	942,310	FL - SE	2
1942	1	250	19.2%	48	NOAA	13,296	TX - No	1
1942	2	26,500	19.2%	5,099	NOAA	1,028,039	TX - No, Ce	3
1943	1	17,000	19.5%	3,308	NOAA	970,828	TX - No	2
1944	3	250	19.7%	49	NOAA	8,796	NC	1
1944	7	100,000	19.7%	19,680	NOAA	2,087,738	MA	3
1944	11	63,000	19.7%	12,398	NOAA	5,855,343	FL -SW	3
1945	1	250	19.9%	50	NOAA	20,416	FL - SW	1
1945	5	20,000	19.9%	3,980	NOAA	825,054	TX - No, Ce	2
1945	9	60,000	19.9%	11,940	NOAA	3,762,550	FL -SE	3
1946	5	5,200	20.1%	1,046	NOAA	465,074	FL - SW	1
1947	3	250	20.3%	51	NOAA	10,278	TX - No	1
1947	4	110,000	20.3%	22,374	NOAA	5,432,151	FL - SE	4
1947	8	23,000	20.3%	4,678	NOAA	1,460,391	FL - SE	2
1948	5	900	20.6%	185	NOAA	17,116	LA	1
1948	7	12,000	20.6%	2,467	NOAA	668,635	FL - SE	3
1948	8	5,500	20.6%	1,131	NOAA	224,907	FL - SE	2
1949	1	250	20.8%	52	NOAA	11,446	NC	1
1949	2			8,300	PCS	2,728,296	FL - SE	3
1949	10	6,700	20.8%	1,392	NOAA	217,219	TX - No	2
1950	Baker	500	21.0%	105	NOAA	13,449	AL	1
1950	Easy	3,300	21.0%	693	NOAA	194,890	FL - SW	3
1950	King			10,386	PCS	2,853,627	FL - SE	3
1952	Able	2,800	22.5%	630	NOAA	55,046	SC	1
1953	Barbara	1,000	23.3%	233	NOAA	19,612	NC	1
1953	Carol	500	23.3%	116	NOAA	63,152	ME	1
1953	Florence	500	23.3%	116	NOAA	10,799	FL - NW	1
1954	Carol			136,000	PCS	6,265,912	MA	3
1954	Edna			11,500	PCS	643,598	MA	3
1954	Hazel			122,000	PCS	8,196,810	NC	4
1955	Connie			25,200	PCS	1,378,549	MD	3
1955	Diane	800,000	24.8%	9,911	NOAA	696,402	NC	1
1955	Ione			4,500	PCS	362,090	NC	3
1956	Flossy			3,700	PCS	275,001	LA	2

Hurricane Loss Estimates
Continental U.S. 1900 - 1999
Dollars in Thousands

Year	Hurricane Number/ Name	Total Estimated Actual Loss at Time of Event: Economic	Insurance Utilization	Insured	Source	Insured Loss Normalized To 2000	Max Loss State/Region	Max Category
1957	Audrey			$ 32,000	PCS	$ 1,176,396	LA	4
1959	Cindy	$ 500	27.8%	139	NOAA	5,717	SC	1
1959	Debra			7,900	PCS	393,073	TX - No	1
1959	Gracie			13,000	PCS	605,316	SC	3
1960	Donna			91,000	PCS	4,709,959	FL - SE	4
1960	Ethel	1,000	28.6%	286	NOAA	11,837	MS	1
1961	Carla			100,000	PCS	3,476,218	TX - No, Ce	4
1963	Cindy			154	PCS	3,954	TX - No	1
1964	Cleo			67,200	PCS	3,746,855	FL - SE	2
1964	Dora			12,000	PCS	403,169	FL - NE	2
1964	Hilda			23,000	PCS	596,026	LA	3
1964	Isabel			2,000	PCS	122,518	FL - SE	2
1965	Betsy			515,000	PCS	11,518,111	LA	3
1966	Alma			5,400	PCS	194,630	FL - SW	2
1966	Inez			596	PCS	16,208	FL - SE	1
1967	Beulah			34,800	PCS	888,088	TX - So	3
1968	Gladys			2,580	PCS	96,877	FL - SW	2
1969	Camille			166,000	PCS	2,949,789	MS	5
1969	Gerda	500	35.4%	177	NOAA	2,439	ME	1
1970	Celia			309,950	PCS	4,568,366	TX - Ce, So	3
1971	Fern			1,380	PCS	18,825	TX - No, Ce	2
1971	Edith			5,730	PCS	71,158	LA	1
1971	Ginger			2,000	PCS	31,447	NC	1
1972	Agnes			101,948	PCS	956,927	PA	1
1974	Carmen			14,721	PCS	118,642	LA	3
1975	Eloise			77,868	PCS	783,072	FL - NW	3
1976	Belle			22,697	PCS	127,951	NY	1
1977	Babe			2,000	PCS	11,414	LA	1
1979	Bob	20,000	42.9%	8,582	NOAA	34,636	LA	1
1979	David			86,990	PCS	547,711	FL - NE	2
1979	Frederic			742,044	PCS	3,686,521	AL	3
1980	Allen			57,611	PCS	283,869	TX - So	3
1983	Alicia			1,274,500	AIRAC	3,912,101	TX - No	3
1984	Diana			36,000	AIRAC	133,682	NC	3
1985	Bob			10,000	AIRAC	29,419	SC	1
1985	Danny			24,500	AIRAC	58,548	LA	1
1985	Elena			622,000	AIRAC	1,650,468	MS	3
1985	Gloria			618,300	AIRAC	1,435,127	NY	3
1985	Juan			78,500	AIRAC	192,283	LA	1
1985	Kate			67,800	AIRAC	189,781	FL -NW	2
1986	Bonnie			21,269	PCS	42,825	TX - No	1
1986	Charley			7,000	PCS	19,357	NC	1
1987	Floyd	500	49.0%	245	NOAA	502	FL - SW	1
1988	Florence			10,000	PCS	19,065	LA	1
1989	Chantal			40,000	PCS	69,972	TX - No	1
1989	Hugo			2,955,000	PCS	5,529,261	SC	4
1989	Jerry			35,000	PCS	63,918	TX - No	1
1991	Bob			610,000	PCS	923,918	MA	2
1992	Andrew			16,600,000	FL Dept	24,486,691	FL - SE	4
1993	Emily			30,000	PCS	47,299	NC	3
1995	Erin			375,000	PCS	484,223	FL - NW, NE	1
1995	Opal			1,990,000	PCS	2,584,891	FL - NW	3

Hurricane Loss Estimates
Continental U.S. 1900 - 1999
Dollars in Thousands

Year	Hurricane Number/ Name	Total Estimated Actual Loss at Time of Event: Economic	Insurance Utilization	Insured	Source	Insured Loss Normalized To 2000	Max Loss State/Region	Max Category
1996	Bertha			$ 135,000	PCS	$ 169,071	NC	2
1996	Fran			1,535,000	PCS	1,910,703	NC	3
1997	Danny			35,000	PCS	41,277	AL	1
1998	Bonnie			360,000	PCS	400,501	NC	2
1998	Earl			18,000	PCS	19,929	FL - NW	1
1998	Georges			1,155,000	PCS	1,270,333	FL - SW	2
1999	Bret			30,000	PCS	31,388	TX - So	3
1999	Floyd			1,875,000	PCS	1,979,274	NC	2

#	%	Category	Insured	Sum	%	Average
62	37.8%	1		7,573,283	2.6%	$ 122,150
38	23.2%	2		24,289,360	8.5%	639,194
47	28.7%	3		93,362,199	32.5%	1,986,430
15	9.1%	4		157,930,884	55.0%	10,528,726
2	1.2%	5		4,141,174	1.4%	2,070,587
164	100.0%	All	33,586,399	287,296,900		1,751,810

#	%	Decade	Sum	%	Average
15	9.1%	Aughts	31,942,476	11.1%	2,129,498
20	12.2%	Teens	36,264,818	12.6%	1,813,241
15	9.1%	Twenties	64,400,759	22.4%	4,293,384
17	10.4%	Thirties	16,689,841	5.8%	981,755
23	14.0%	Forties	27,116,547	9.4%	1,178,980
18	11.0%	Fifties	23,209,438	8.1%	1,289,413
15	9.1%	Sixties	28,736,676	10.0%	1,915,778
12	7.3%	Seventies	10,956,670	3.8%	913,056
16	9.8%	Eighties	13,630,178	4.7%	851,886
13	7.9%	Nineties	34,349,498	12.0%	2,642,269
164		All	287,296,900		1,751,810

Notes:

Where based on NOAA, insured loss equals economic loss times insurance utilization factor times flood adjustment factor. Only the following storms, which had unusual amounts of uninsured flood damage, were reduced to reflect flood: 1900 #1 (50%), 1915 #2 (75%), 1916 #1 (50%), 1955 Diane (5%).

Economic losses for smaller events estimated judgmentally.

PCS losses exclude the following states and territories, which were excluded from the normalization model:

1975 Eloise	PA, PR
1979 David	PR, VI, VA to MA
1979 Frederic	KY, NY, OH, PA, WV
1980 Allen	PR, VI
1989 Hugo	PR, VI
1995 Opal	NC, SC, TN
1996 Fran	PA, OH
1997 Danny	NC, SC
1998 Georges	PR, VI
1999 Floyd	PA, RI

Normalized Hurricane Loss - Annual Aggregate Severity Distributions by State
Dollars in Thousands

State	Normalized Actual 20th Century Return Period (Years) 100	50	25	20	10	5	Expected Annual	% of Total
Texas	$ 16,357,807	$ 16,044,802	$ 4,568,366	$ 3,912,101	$ 959,320	$ 133,890	$ 615,179	21.4%
Louisiana	10,426,919	1,642,437	1,115,135	723,002	343,527	30,640	195,641	6.8%
Mississippi	2,490,730	1,337,271	799,333	735,718	159,861	3,683	77,431	2.7%
Alabama	2,406,881	1,363,217	385,039	379,566	31,137	1,335	61,380	2.1%
Florida	49,744,060	23,763,689	7,976,601	5,837,485	3,052,795	910,060	1,422,764	49.5%
Georgia	429,105	176,122	101,460	73,375	15,783	1,094	11,487	0.4%
South Carolina	4,140,037	606,128	244,375	220,535	40,168	5,947	61,660	2.1%
North Carolina	1,943,528	1,768,044	1,399,847	1,371,862	267,909	23,152	109,399	3.8%
Virginia	2,188,909	872,795	112,753	104,579	33,871	842	38,253	1.3%
Maryland	834,038	484,365	53,170	48,076	5,340	-	16,951	0.6%
Delaware	341,019	26,476	14,979	14,200	365	-	4,360	0.2%
New Jersey	980,301	600,714	99,297	92,297	32,234	-	22,166	0.8%
New York	3,082,156	1,490,510	208,076	183,374	36,439	-	61,227	2.1%
Connecticut	4,095,213	504,385	151,939	76,484	50	-	50,944	1.8%
Rhode Island	1,322,697	416,528	160,166	134,081	-	-	24,819	0.9%
Massachusetts	2,904,903	1,484,027	456,272	367,780	924	-	63,812	2.2%
New Hampshire	412,611	159,311	11,635	10,464	-	-	6,178	0.2%
Maine	285,940	56,837	18,511	17,402	-	-	4,175	0.1%
Total All States	51,789,586	24,486,691	16,485,683	15,106,320	9,373,159	3,555,627	2,872,969	

Note: Return period loss based on distribution by state of normalized losses in Exhibit 3, e.g., 100 year return is the worst year in the 20th century, 50 year return is the second worst year, 25 year return is the 4th worst year, etc. Not to be confused with probabilistic return period distributions and expected losses based on catastrophe models, which are intended to reflect longer term probabilities.

Normalized Hurricane Loss - Maximum Single Occurrence Severity Distributions by State
Dollars in Thousands

State	Normalized Actual 20th Century Return Period (Years)						100 Year Event
	100	50	25	20	10	5	
Texas	$ 16,357,807	$ 16,044,802	$ 4,568,366	$ 3,912,101	$ 959,320	$ 69,972	1900 - 1 ("Isaac's")
Louisiana	10,426,919	1,540,864	1,115,135	723,002	343,527	28,513	1965 - Betsy
Mississippi	2,490,730	1,331,575	793,954	735,718	159,861	3,683	1969 - Camille
Alabama	2,406,138	1,363,217	353,807	218,189	31,137	1,335	1979 - Frederick
Florida	47,989,146	23,763,689	7,976,601	5,837,485	2,853,627	894,836	1926 - 6
Georgia	429,105	176,122	101,460	73,375	15,783	1,094	1995 - Opal
South Carolina	4,140,037	605,316	244,375	220,535	37,008	5,947	1989 - Hugo
North Carolina	1,943,528	1,641,766	1,371,862	641,628	267,909	23,152	1954 - Hazel
Virginia	2,188,909	854,007	112,753	104,579	33,871	842	1954 - Hazel
Maryland	834,038	484,106	53,170	48,076	5,340	-	1954 - Hazel
Delaware	341,019	26,476	14,979	14,200	365	-	1954 - Hazel
New Jersey	600,714	579,055	99,297	92,297	32,234	-	1938 - 4 or 1954 - Hazel
New York	3,082,156	1,077,727	208,076	183,374	36,439	-	1938 - 4 ("Great New England")
Connecticut	4,095,213	351,008	151,939	76,484	50	-	1938 - 4 ("Great New England")
Rhode Island	1,183,942	416,528	160,166	134,081	-	-	1954 - Carol
Massachusetts	2,655,844	1,484,027	456,272	367,780	924	-	1954 - Carol
New Hampshire	332,968	159,311	11,635	10,464	-	-	1954 - Carol
Maine	263,178	56,837	18,511	17,402	-	-	1954 - Carol
Total All States	49,728,840	24,486,691	16,146,375	11,518,111	7,976,601	3,476,218	

Note: Return period loss based on distribution by state of the largest normalized loss per year in Exhibit 3, e.g., 100 year return is the worst event, 50 year return is the second worst event, 25 year return is the 4th worst event, etc. Not to be confused with probabilistic return period distributions and expected losses based on catastrophe models, which are intended to reflect longer term probabilities.

Normalized Hurricane Loss - Annual Aggregate Severity Distributions by State and County
Counties with Significant Annual Expected Losses
Dollars in Thousands

State	County	Estimated 2000 Housing Units	Normalized Actual 20th Century Return Period (Years) 100	50	25	20	10	5	Expected Annual	Expected Loss Per Unit ($'s)
TX										
	Harris	1,305,351	$9,953,674	$8,841,048	$729,077	$560,265	$199,602	$0	$245,595	$188
	Galveston	110,157	4,506,461	4,084,453	360,805	315,733	44,502	1,106	104,432	948
	Nueces	122,333	7,287,137	2,001,912	90,356	53,950	36,982	0	98,660	806
	Brazoria	88,261	1,359,509	581,793	166,175	164,674	28,757	434	33,046	374
	Fort Bend	121,367	911,594	401,431	160,493	153,787	14,463	0	23,965	197
	Cameron	114,432	647,510	513,497	68,357	32,195	3,878	0	14,581	127
	Aransas	14,188	1,203,723	114,140	6,721	4,802	1,624	0	14,044	990
	San Patricio	26,640	1,032,527	136,714	6,968	5,220	3,636	0	12,619	474
	Montgomery	114,584	285,815	244,840	53,763	22,594	3,909	0	7,953	69
	Hidalgo	184,668	555,041	119,872	14,585	5,975	0	0	7,720	42
	Jefferson	97,558	261,334	165,980	33,504	21,097	7,430	32	6,103	63
	Matagorda	18,329	179,112	141,720	42,226	9,220	1,892	206	4,539	248
	Chambers	9,305	145,296	127,939	8,388	4,940	1,430	11	3,147	338
	Victoria	31,792	268,874	14,153	5,013	1,338	355	0	3,067	96
FL										
	Dade	860,587	24,841,690	21,503,754	2,448,916	1,154,922	528,163	32,634	594,201	690
	Broward	784,873	8,274,310	1,837,931	1,275,267	1,250,347	432,580	30,674	188,435	240
	Palm Beach	580,029	2,613,939	2,449,415	1,278,092	874,908	186,600	30,599	119,848	207
	Monroe	48,610	3,285,189	1,306,132	815,359	659,162	93,993	8,586	86,746	1,785
	Lee	232,004	4,333,589	1,174,856	282,775	278,928	47,403	14,434	75,937	327
	Escambia	122,238	1,242,614	537,338	243,515	86,124	8,999	156	26,799	219
	Brevard	228,560	805,310	688,639	202,758	173,427	23,231	2,025	25,084	110
	Collier	134,052	1,510,837	345,577	110,492	68,745	12,317	4,488	25,022	187
	Sarasota	174,066	1,157,395	723,028	112,817	51,022	23,990	5,187	24,846	143
	Pinellas	470,889	603,486	470,479	152,418	95,421	58,754	9,286	23,269	49
	Santa Rosa	52,623	961,706	639,907	150,955	83,197	8,161	250	22,866	435
	St. Lucie	94,666	1,110,664	376,664	115,185	76,408	24,309	1,799	21,996	232
	Hillsborough	413,122	749,675	222,368	134,788	95,736	26,053	4,100	16,790	41
	Okaloosa	79,064	632,113	336,647	121,265	60,763	5,794	336	14,755	187
	Martin	64,667	619,485	272,745	117,000	74,602	10,303	1,420	14,627	226
	Manatee	133,772	483,954	468,404	71,797	39,658	23,189	3,284	13,879	104
	Volusia	216,688	314,543	278,535	148,743	137,068	14,648	1,118	13,635	63
	Orange	339,869	411,441	196,923	134,578	122,244	20,628	343	13,610	40
	Polk	213,034	375,193	365,058	124,589	109,153	16,023	1,041	13,420	63
	Indian River	52,411	562,726	174,576	40,527	37,628	12,896	509	11,084	211
	Charlotte	84,296	568,944	270,309	22,544	16,893	6,665	761	10,038	119
	Pasco	175,854	219,943	162,509	47,060	30,902	11,942	1,696	6,880	39
	Lake	106,250	186,706	179,379	44,272	42,158	8,788	301	6,538	62
	Seminole	152,097	145,588	95,484	61,216	55,408	6,372	0	5,571	37
	Duval	317,548	232,279	84,687	46,432	28,001	5,734	0	5,544	17
	Bay	81,598	264,066	100,921	36,975	17,167	5,810	0	5,423	66
	Osceola	70,504	148,485	65,616	36,843	23,752	6,872	219	4,080	58
	Marion	124,315	131,971	107,128	23,137	22,395	8,516	243	4,071	33
	Highlands	46,304	60,603	52,898	25,421	22,991	2,042	236	2,745	59

Note: Return period loss based on distribution by state and county of normalized losses in Exhibit 3, e.g., 100 year return is the worst year in the 20th century, 50 year return is the second worst year, 25 year return is the 4th worst year, etc. Not to be confused with probabilistic return period distributions and expected losses based on catastrophe models, which are intended to reflect longer term probabilities. Expected loss per unit compares expected annual losses (personal, commercial, and auto) with residential - only housing units, i.e., it is intended as a relative measure of cost per unit of exposure but not as a measure of residential costs per unit.

Comparison of Actual vs. Modeled Hurricane Experience

Number of Landfalling Storms per Century

Category	Actual 20th Century CW	TX	FL	Model T CW	TX	FL
1	63	12	17	62.0	11.0	16.5
2	36	8	15	37.5	8.5	15.0
3	48	10	16	46.0	9.5	17.0
4	15	6	5	16.0	5.0	6.0
5	2	0	1	2.5	0.5	1.0
All	164	36	54	164.0	34.5	55.5

Estimated Annual Aggregate Insured Loss ($000)

Category	Normalized 20th Century CW	TX	FL	Model T CW	TX	FL
1	$ 75,733	$ 9,146	$ 44,648	$ 59,199	$ 5,176	$ 26,473
2	242,894	22,175	134,857	300,721	34,207	143,086
3	933,622	131,962	329,344	852,477	88,322	391,428
4	1,579,309	451,897	902,001	1,262,920	186,123	714,092
5	41,412	-	11,914	403,634	65,421	191,347
Expected	2,872,969	615,179	1,422,764	2,878,951	379,250	1,466,427

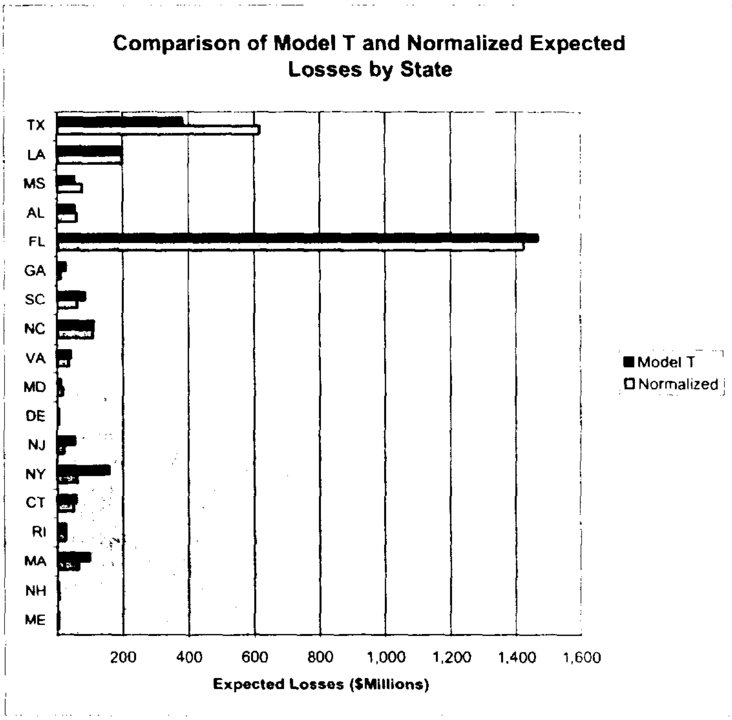
Estimated Annual Aggregate Insured Loss ($000)

Return Period (Yrs)	Normalized 20th Century CW	TX	FL	Model T CW	TX	FL
5	$ 3,555,627	$ 133,890	$ 910,060	$ 3,569,742	$ 126,796	$ 954,030
10	9,373,159	959,320	3,052,795	6,917,383	684,396	3,206,555
20	15,106,320	3,912,101	5,837,485	11,780,896	2,032,334	7,702,533
25	16,485,683	4,568,366	7,976,601	14,687,232	2,821,885	10,343,645
50	24,486,691	16,044,802	23,763,689	21,710,120	5,061,653	17,296,870
100	51,789,586	16,357,807	49,744,060	33,133,590	8,331,148	28,926,913
Expected	2,872,969	615,179	1,422,764	2,878,951	379,250	1,466,427

Notes: Countrywide (CW) normalized figures based on continental U.S. from Exhibits 3 and 4.
Texas and Florida actual frequencies from Exhibit 1.
Texas and Florida normalized damages from Exhibit 4 and underlying data.
Model T is a hypothetical probabilistic hurricane model

Comparison of Actual vs. Modeled Hurricane Expected Losses by State

	Annual Expected Losses ($000)		
State	Normalized Actual	Model T	Model T Difference
Texas	$ 615,179	$ 379,250	-38%
Louisiana	195,641	197,501	1%
Mississippi	77,431	54,460	-30%
Alabama	61,380	54,522	-11%
Florida	1,422,764	1,466,427	3%
Georgia	11,487	27,849	142%
South Carolina	61,660	84,864	38%
North Carolina	109,399	110,872	1%
Virginia	38,253	43,274	13%
Maryland	16,951	11,685	-31%
Delaware	4,360	2,766	-37%
New Jersey	22,166	52,633	137%
New York	61,227	157,509	157%
Connecticut	50,944	59,280	16%
Rhode Island	24,819	26,220	6%
Massachusetts	63,812	96,552	51%
New Hampshire	6,178	4,721	-24%
Maine	4,175	4,830	16%
All States	2,872,969	2,878,951	0%



Notes: Normalized figures from Exhibit 4, Sheet 1
Model T is a hypothetical probabilistic hurricane model

Comparison of Actual vs. Modeled Hurricane Losses
Top 50 Historical Normalized Events

Rank	Year	Number/ Name	Normalized	Model T	Max Loss State/Region	Max Category
1	1926	6	49,728,840	44,000,000	FL - SE	4
2	1992	Andrew	24,486,691	24,900,000	FL - SE	4
3	1900	1	16,485,683	11,900,000	TX - No	4
4	1915	2	16,146,375	9,800,000	TX - No	4
5	1965	Betsy	11,518,111	12,900,000	LA	3
6	1919	2	10,009,409	4,800,000	FL - SW	4
7	1938	4	9,965,606	12,800,000	CT	3
8	1928	4	9,816,472	16,700,000	FL - SE	4
9	1954	Hazel	8,196,810	6,700,000	NC	4
10	1909	9	7,976,601	3,400,000	FL - SE	3
11	1954	Carol	6,265,912	5,600,000	MA	3
12	1944	11	5,855,343	9,700,000	FL -SW	3
13	1989	Hugo	5,529,261	5,900,000	SC	4
14	1947	4	5,432,151	17,600,000	FL - SE	4
15	1960	Donna	4,709,959	8,800,000	FL - SE	4
16	1970	Celia	4,568,366	4,400,000	TX - Ce, So	3
17	1983	Alicia	3,912,101	2,800,000	TX - No	3
18	1945	9	3,762,550	6,600,000	FL -SE	3
19	1964	Cleo	3,746,855	2,900,000	FL - SE	2
20	1979	Frederic	3,686,521	2,100,000	AL	3
21	1961	Carla	3,476,218	2,600,000	TX - No, Ce	4
22	1916	1	3,096,434	2,300,000	MS	3
23	1969	Camille	2,949,789	3,300,000	MS	5
24	1950	King	2,853,627	7,500,000	FL - SE	3
25	1910	4	2,735,157	3,100,000	FL - SW	3
26	1949	2	2,728,296	6,700,000	FL - SE	3
27	1995	Opal	2,584,891	2,400,000	FL - NW	3
28	1903	3	2,124,106	2,600,000	FL - SE	2
29	1944	7	2,087,738	4,500,000	MA	3
30	1999	Floyd	1,979,274	2,000,000	NC	2
31	1996	Fran	1,910,703	2,100,000	NC	3
32	1926	1	1,755,434	2,700,000	FL - NE	2
33	1915	5	1,709,809	2,700,000	LA	4
34	1985	Elena	1,650,468	1,300,000	MS	3
35	1921	6	1,624,995	5,400,000	FL - SW	3
36	1947	8	1,460,391	1,200,000	FL - SE	2
37	1985	Gloria	1,435,127	1,900,000	NY	3
38	1955	Connie	1,378,549	1,700,000	MD	3
39	1935	6	1,371,030	1,500,000	FL - SE	2
40	1933	8	1,356,989	1,300,000	VA	2
41	1998	Georges	1,270,333	1,300,000	FL - SW	2
42	1935	2	1,191,386	2,400,000	FL - SW	5
43	1957	Audrey	1,176,396	1,000,000	LA	4
44	1933	12	1,163,819	3,900,000	FL -SE	3
45	1909	3	1,119,560	1,600,000	TX - No	3
46	1942	2	1,028,039	500,000	TX - No, Ce	3
47	1943	1	970,828	700,000	TX - No	2
48	1972	Agnes	956,927	400,000	PA	1
49	1941	5	942,310	8,100,000	FL - SE	2
50	1991	Bob	923,918	1,300,000	MA	2
			264,812,155	294,300,000		

Notes: Normalized figures from Exhibit 3
Model T is a hypothetical hurricane model

Hurricanes Affecting the Bermuda, Hawaii, Puerto Rico and USVI 1900-1999

		Date of	Category and Key Islands Affected								
	Number/	First		Hawaiian Islands			Puerto	US Virgin Islands		PR or	US Landfall
Year	Name	Landfall	Bermuda	Hawaii	Kauai	Oahu	Rico	St. Thomas	St. Croix	USVI	States Affected and Category
1900	4	17-Sep	1								None
1903	6	28-Sep	1								None
1915	3	03-Sep	1								None
1916	10	23-Sep	1								None
1918	4	04-Sep	1								None
1921	3	15-Sep	1								None
1922	2	21-Sep	2								None
1926	10	22-Oct	3								None
1939	4	16-Oct	3								None
1947	9	20-Oct	2								None
1948	6	13-Sep	2								None
1948	8	07-Oct	2								FLSE 2
1953	Edna	17-Sep	2								None
1963	Arlene	09-Aug	1								None
1987	Emily	24-Sep	2								None
1989	Dean	06-Aug	1								None
1999	Gert	21-Sep	1								None
1950	Hiki	15-Aug			1						
1957	Nina	02-Dec			1						
1959	Dot	06-Aug			2						
1982	Iwa	23-Nov			1	1					
1992	Iniki	11-Sep			4						
1916	5/San Hipolito	22 Aug					1	2	2	2	None
1916	12							2	1	2	None
1926	1/San Liborio	23-Jul					1			1	FLNE 2
1928	4/San Felipe	13-Sep					5		5	5	FLSE 4, FLNE 2 GA 1, SC 1
1930	2	02-Sep					1			1	None
1931	6/San Nicolas	10-Sep					2	2	1	2	None
1932	7/San Ciprian	26-Sep					2	2	1	2	None
1956	Santo Clara (Betsy)	12-Aug					1			1	None
1960	Donna	05-Sep						1		1	FLSW 4, NC 3, NY 3
1989	Hugo	18-Sep					4	3	4	4	SC 4
1995	Marilyn	16-Sep						2	2	2	None
1996	Bertha	08-Jul						1		1	NC 2
1996	Hortense	10-Sep					1			1	None
1998	Georges	21-Sep					2	1	2	2	FLSW 2, MS 2
1999	Lenny	17-Nov							1	1	None
	Category 1		9	0	3	1	5	3	4	7	
	Category 2		6	0	1	0	3	5	3	6	
	Category 3		2	0	0	0	0	1	0	0	
	Category 4		0	0	1	0	1	0	1	1	
	Category 5		0	0	0	0	1	0	1	1	
	Total		17	0	5	1	10	9	9	15	

Note: Category designations, according to Saffir/Simpson Hurricane Scale, based on estimated sustained winds over land reflecting authors' judgment based on review of:

- NOAA summary reports and best track files (www.nhc.noaa.gov/pastall.html)
- Neumann (Newmann, Jarvinen, McAdie and Elms, 1993, p. 31)
- Hebert (Hebert, Jarrell and Mayfield, 1996, Table 14)
- Tucker (Tucker, 1995)

No hurricanes have affected the west coast of the U.S. during the 20th century. According to the National Weather Service office in Oxnard, California, two storms are recognized as having produced tropical storm conditions over land:

- September 25, 1939 in Southern California (Long Beach area)
- October 6, 1972 in Arizona (remnants of Hurricane Joanne)

Hurricanes Affecting the Bermuda, Hawaii, Puerto Rico and USVI 1900-1999
Estimated Damage at Time of Event
Dollars in Thousands

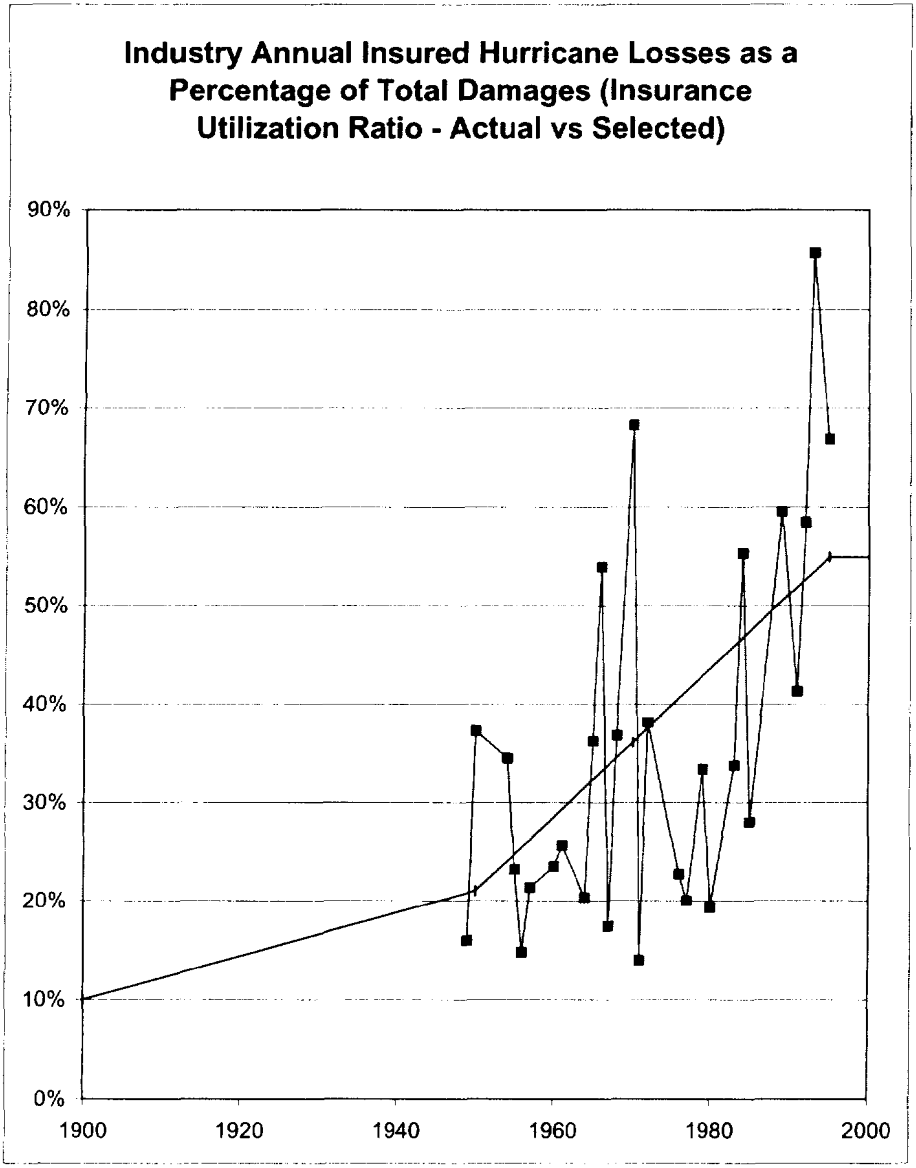
Year	Number/ Name	Estimated Damage Economic	Estimated Damage Insured	Source
Bermuda				
1900	4	Unk		
1903	6	Unk		
1915	3	Unk		
1916	10	Unk		
1918	4	Unk		
1921	3	Unk		
1922	2	Unk		
1926	10	Unk		
1939	4	Unk		
1947	9	Unk		
1948	6	Unk		
1948	8	Unk		
1953	Edna	Unk		
1963	Arlene		75	Tucker
1987	Emily		35,000	NOAA
1989	Dean		5,000	NOAA
1999	Gert		Unk	
Hawaii				
1950	Hiki	Unk		
1957	Nina	200		Hebert
1959	Dot	6,000		Hebert
1982	Iwa		137,000	PCS
1992	Iniki		1,906,000	PCS

Year	Number/ Name	Economic	Insured PR	Insured USVI	Source
Puerto Rico and USVI					
1916	5/San Hipolito	1,000			Hebert
1916	12	Unk			
1926	1/San Liborio	5,000			Hebert
1928	4/San Felipe	85,000			Hebert
1930	2	Unk			
1931	6/San Nicolas	200			Hebert
1932	7/San Ciprian	30,000			Hebert
1956	Santo Clara (Betsy)	40,000	10,000		PCS
1960	Donna	Unk			Hebert
1989	Hugo		440,000	800,000	PCS
1995	Marilyn		75,000	800,000	PCS
1996	Bertha			Unk	
1996	Hortense		150,000		PCS
1998	Georges		1,750,000	50,000	PCS
1999	Lenny			Unk	

Historical Indices Used in Normalization Model
Annual Growth Rates

Year	Implicit Price Deflator	Net Stock of FRTW	National Housing Units	National Population	Insurance Utilization
1901	3.5%	2.5%	1.9%	1.9%	2.2%
1902	3.5%	2.5%	1.9%	1.9%	2.2%
1903	3.5%	2.5%	1.9%	1.9%	2.1%
1904	3.5%	2.5%	1.9%	1.9%	2.1%
1905	3.5%	2.5%	1.9%	1.9%	2.0%
1906	3.5%	2.5%	1.9%	1.9%	2.0%
1907	3.5%	2.5%	1.9%	1.9%	1.9%
1908	3.5%	2.5%	1.9%	1.9%	1.9%
1909	3.5%	2.5%	1.9%	1.9%	1.9%
1910	3.5%	2.5%	1.9%	1.9%	1.8%
1911	3.5%	2.5%	1.4%	1.4%	1.8%
1912	3.5%	2.5%	1.4%	1.4%	1.8%
1913	3.5%	2.5%	1.4%	1.4%	1.7%
1914	3.5%	2.5%	1.4%	1.4%	1.7%
1915	3.5%	2.5%	1.4%	1.4%	1.7%
1916	3.5%	2.5%	1.4%	1.4%	1.7%
1917	3.5%	2.5%	1.4%	1.4%	1.6%
1918	3.5%	2.5%	1.4%	1.4%	1.6%
1919	3.5%	2.5%	1.4%	1.4%	1.6%
1920	3.5%	2.5%	1.4%	1.4%	1.6%
1921	3.5%	2.5%	1.5%	1.5%	1.5%
1922	3.5%	2.5%	1.5%	1.5%	1.5%
1923	3.5%	2.5%	1.5%	1.5%	1.5%
1924	3.5%	2.5%	1.5%	1.5%	1.5%
1925	3.5%	2.5%	1.5%	1.5%	1.4%
1926	3.5%	4.1%	1.5%	1.5%	1.4%
1927	3.5%	3.6%	1.5%	1.5%	1.4%
1928	3.5%	3.2%	1.5%	1.5%	1.4%
1929	3.5%	3.2%	1.5%	1.5%	1.4%
1930	3.5%	1.7%	1.5%	1.5%	1.3%
1931	3.5%	0.4%	0.7%	0.7%	1.3%
1932	3.5%	-1.0%	0.7%	0.7%	1.3%
1933	3.5%	-1.3%	0.7%	0.7%	1.3%
1934	3.5%	-0.6%	0.7%	0.7%	1.3%
1935	3.5%	0.2%	0.7%	0.7%	1.3%
1936	3.5%	1.5%	0.7%	0.7%	1.2%
1937	3.5%	1.8%	0.7%	0.7%	1.2%
1938	3.5%	0.9%	0.7%	0.7%	1.2%
1939	3.5%	1.7%	0.7%	0.7%	1.2%
1940	3.5%	2.1%	0.7%	0.7%	1.2%
1941	3.5%	3.7%	2.1%	1.4%	1.2%
1942	3.5%	5.4%	2.1%	1.4%	1.2%
1943	3.5%	5.8%	2.1%	1.4%	1.1%
1944	3.5%	4.6%	2.1%	1.4%	1.1%
1945	3.5%	2.1%	2.1%	1.4%	1.1%
1946	3.5%	0.4%	2.1%	1.4%	1.1%
1947	3.5%	1.4%	2.1%	1.4%	1.1%
1948	3.5%	2.1%	2.1%	1.4%	1.1%
1949	3.5%	2.6%	2.1%	1.4%	1.1%
1950	3.5%	3.7%	2.1%	1.4%	1.1%
1951	5.5%	4.0%	2.4%	1.7%	3.6%
1952	1.4%	3.8%	2.4%	1.7%	3.5%
1953	0.9%	4.2%	2.4%	1.7%	3.4%
1954	0.9%	3.7%	2.4%	1.7%	3.2%
1955	2.7%	4.3%	2.4%	1.7%	3.1%
1956	3.2%	3.7%	2.4%	1.7%	3.0%
1957	2.8%	3.4%	2.4%	1.7%	3.0%
1958	2.7%	2.8%	2.4%	1.7%	2.9%
1959	0.8%	3.6%	2.4%	1.7%	2.8%
1960	1.6%	3.3%	2.4%	1.7%	2.7%
1961	1.0%	3.1%	1.6%	1.3%	2.6%
1962	1.3%	3.5%	1.6%	1.3%	2.6%
1963	1.5%	3.7%	1.6%	1.3%	2.5%
1964	1.4%	4.1%	1.6%	1.3%	2.5%
1965	2.2%	4.4%	1.6%	1.3%	2.4%
1966	3.4%	4.5%	1.6%	1.3%	2.3%
1967	3.4%	4.0%	1.6%	1.3%	2.3%
1968	4.5%	4.1%	1.6%	1.3%	2.2%
1969	4.9%	3.9%	1.6%	1.3%	2.2%
1970	5.1%	3.2%	1.6%	1.3%	2.1%
1971	4.9%	3.3%	2.6%	1.1%	2.1%
1972	4.4%	4.0%	2.6%	1.1%	2.0%
1973	6.9%	3.9%	2.6%	1.1%	2.0%
1974	10.6%	3.0%	2.6%	1.1%	2.0%
1975	7.6%	2.2%	2.6%	1.1%	1.9%
1976	5.5%	2.6%	2.6%	1.1%	1.9%
1977	6.7%	3.1%	2.6%	1.1%	1.9%
1978	7.7%	3.5%	2.6%	1.1%	1.8%
1979	8.7%	3.4%	2.6%	1.1%	1.8%
1980	10.0%	2.5%	2.6%	1.1%	1.8%
1981	8.4%	2.4%	1.5%	0.9%	1.7%
1982	5.2%	1.8%	1.5%	0.9%	1.7%
1983	3.9%	2.2%	1.5%	0.9%	1.7%
1984	3.5%	3.1%	1.5%	0.9%	1.6%
1985	3.4%	3.3%	1.5%	0.9%	1.6%
1986	2.5%	2.2%	1.5%	0.9%	1.6%
1987	3.2%	3.0%	1.5%	0.9%	1.6%
1988	4.0%	2.9%	1.5%	0.9%	1.5%
1989	3.9%	2.6%	1.5%	0.9%	1.5%
1990	4.6%	2.3%	1.5%	0.9%	1.5%
1991	3.4%	1.6%	1.2%	1.0%	1.5%
1992	2.6%	1.7%	1.2%	1.0%	1.5%
1993	2.6%	2.0%	1.2%	1.0%	1.4%
1994	2.5%	2.2%	1.2%	1.0%	1.4%
1995	2.1%	2.5%	1.2%	1.0%	1.4%
1996	1.8%	2.7%	1.2%	1.0%	0.0%
1997	1.7%	2.7%	1.2%	1.0%	0.0%
1998	1.2%	2.7%	1.2%	1.0%	0.0%
1999	1.5%	2.7%	1.2%	1.0%	0.0%
2000	2.0%	2.7%	1.2%	1.0%	0.0%

Notes: Implicit price deflator available back to 1950; 3.5% trend assumed for 1950 and prior
FRTW is fixed reproducable tangible wealth, Department of Commerce, Bureau of Economic Analysis
- Available back to 1925, 2.5% trend assumed for 1925 and prior
Housing units and population growth based on annual growth between each decennial census
Insurance utilization index based on linear trends from 1900 to 1950 and from 1950 to 1995
- See text and graph on Appendix B, Exhibit 2 for further information


Industry Annual Insured Hurricane Losses as a Percentage of Total Damages (Insurance Utilization Ratio - Actual vs Selected)
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1900
1920
1940
1960
1980
2000

Neural Networks Demystified

Louise Francis, FCAS, MAAA

Title: Neural Networks Demystified
by Louise Francis

Francis Analytics and Actuarial Data Mining, Inc.

Abstract:
This paper will introduce the neural network technique of analyzing data as a generalization of more familiar linear models such as linear regression. The reader is introduced to the traditional explanation of neural networks as being modeled on the functioning of neurons in the brain. Then a comparison is made of the structure and function of neural networks to that of linear models that the reader is more familiar with. The paper will then show that backpropagation neural networks with a single hidden layer are universal function approximators. The paper will also compare neural networks to procedures such as Factor Analysis which perform dimension reduction. The application of both the neural network method and classical statistical procedures to insurance problems such as the prediction of frequencies and severities is illustrated.

One key criticism of neural networks is that they are a "black box". Data goes into the "black box" and a prediction comes out of it, but the nature of the relationship between independent and dependent variables is usually not revealed.. Several methods for interpreting the results of a neural network analysis, including a procedure for visualizing the form of the fitted function will be presented.

Acknowledgments:
The author wishes to acknowledge the following people who reviewed this paper and provided many constructive suggestions: Patricia Francis-Lyon, Virginia Lambert, Francis Murphy and Christopher Yaure

Neural Networks Demystified

Introduction

Artificial neural networks are the intriguing new high tech tool for finding hidden gems in data. They belong to a broader category of techniques for analyzing data known as data mining. Other widely used tools include decision trees, genetic algorithms, regression splines and clustering. Data mining techniques are used to find patterns in data. Typically the data sets are large, i.e. have many records and many predictor variables. The number of records is typically at least in the tens of thousands and the number of independent variables is often in the hundreds. Data mining techniques, including neural networks, have been applied to portfolio selection, credit scoring, fraud detection and market research. When data mining tools are presented with data containing complex relationships they can be trained to identify the relationships. An advantage they have over classical statistical models used to analyze data, such as regression and ANOVA, is that they can fit data where the relation between independent and dependent variables is nonlinear and where the specific form of the nonlinear relationship is unknown.

Artificial neural networks (hereafter referred to as neural networks) share the advantages just described with the many other data mining tools. However, neural networks have a longer history of research and application. As a result, their value in modeling data has been more extensively studied and better established in the literature (Potts, 2000). Moreover, sometimes they have advantages over other data mining tools. For instance, decisions trees, a method of splitting data into homogenous clusters with similar expected values for the dependent variable, are often less effective when the predictor variables are continuous than when they are categorical.[1] Neural networks work well with both categorical and continuous variables.

Neural Networks are among the more glamorous of the data mining techniques. They originated in the artificial intelligence discipline where they are often portrayed as a brain in a computer. Neural networks are designed to incorporate key features of neurons in the brain and to process data in a manner analogous to the human brain. Much of the terminology used to describe and explain neural networks is borrowed from biology. Many other data mining techniques, such as decision trees and regression splines were developed by statisticians and are described in the literature as computationally intensive generalizations of classical linear models. Classical linear models assume that the functional relationship between the independent variables and the dependent variable is linear. Classical modeling also allows linear relationship that result from a transformation of dependent or independent variables, so some nonlinear relationships can be approximated. Neural networks and other data mining techniques do not require that the relationships between predictor and dependent variables be linear (whether or not the variables are transformed).

---

[1] Salford System's course on Advanced CART, October 15, 1999.

The various data mining tools differ in their approach to approximating nonlinear functions and complex data structures. Neural networks use a series of neurons in what is known as the hidden layer that apply nonlinear activation functions to approximate complex functions in the data. The details are discussed in the body of this paper. As the focus of this paper is neural networks, the other data mining techniques will not be discussed further.

Despite their advantages, many statisticians and actuaries are reluctant to embrace neural networks. One reason is that they are a "black box". Because of the complexity of the functions used in the neural network approximations, neural network software typically does not supply the user with information about the nature of the relationship between predictor and target variables. The output of a neural network is a predicted value and some goodness of fit statistics. However, the functional form of the relationship between independent and dependent variables is not made explicit. In addition, the strength of the relationship between dependent and independent variables, i.e., the importance of each variable, is also often not revealed. Classical models as well as other popular data mining techniques, such as decision trees, supply the user with a functional description or map of the relationships.

This paper seeks to open that black box and show what is happening inside the neural networks. While some of the artificial intelligence terminology and description of neural networks will be presented, this paper's approach is predominantly from the statistical perspective. The similarity between neural networks and regression will be shown. This paper will compare and contrast how neural networks and classical modeling techniques deal with three specific modeling challenges: 1) nonlinear functions, 2) correlated data and 3) interactions. How the output of neural networks can be used to better understand the relationships in the data will then be demonstrated.

Types of Neural Networks

A number of different kinds of neural networks exist. This paper will discuss feedforward neural networks with one hidden layer. A feedforward neural network is a network where the signal is passed from an input layer of neurons through a hidden layer to an output layer of neurons. The function of the hidden layer is to process the information from the input layer. The hidden layer is denoted as hidden because it contains neither input nor output data and the output of the hidden layer generally remains unknown to the user. A feedforward neural network can have more than one hidden layer. However such networks are not common. The feedforward network with one hidden layer is one of the most popular kinds of neural networks. It is historically one of the older neural network techniques. As a result, its effectiveness has been established and software for applying it is widely available. The feedforward neural network discussed in this paper is known as a Multilayer Perceptron (MLP). The MLP is a feedforward network which uses supervised learning. The other popular kinds of feedforward networks often incorporate unsupervised learning into the training. A network that is trained using supervised learning is presented with a target variable and fits a function which can be used to predict the target variable. Alternatively, it may classify records into levels of the target variable when the target variable is categorical.

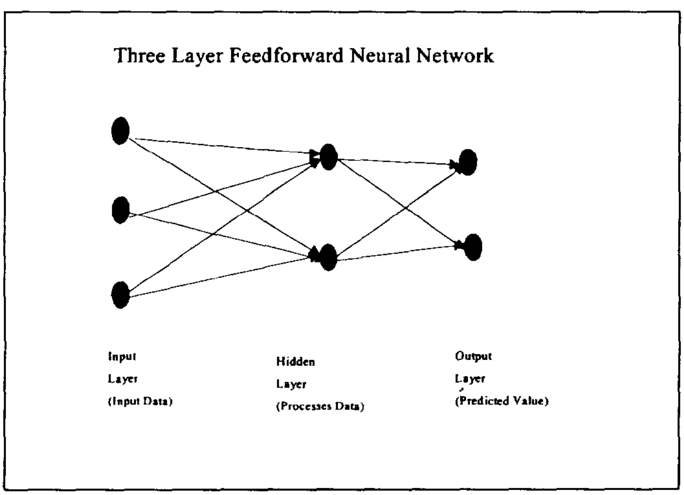
This is analogous to the use of such statistical procedures as regression and logistic regression for prediction and classification. A network trained using unsupervised learning does not have a target variable. The network finds characteristics in the data, which can be used to group similar records together. This is analogous to cluster analysis in classical statistics. This paper will discuss only the former kind of network, and the discussion will be limited to a feedforward MLP neural network with one hidden layer. This paper will primarily present applications of this model to continuous rather than discrete data, but the latter application will also be discussed.

Structure of a Feedforward Neural Network

Figure 1 displays the structure of a feedforward neural network with one hidden layer. The first layer contains the input nodes. Input nodes represent the actual data used to fit a model to the dependent variable and each node is a separate independent variable. These are connected to another layer of neurons called the hidden layer or hidden nodes, which modifies the data. The nodes in the hidden layer connect to the output layer. The output layer represents the target or dependent variable(s). It is common for networks to have only one target variable, or output node, but there can be more. An example would be a classification problem where the target variable can fall into one of a number of categories. Sometimes each of the categories is represented as a separate output node.

As can be seen from the Figure 1, each node in the input layer connects to each node in the hidden layer and each node in the hidden layer connects to each node in the output layer.

Figure 1



This structure is viewed in the artificial intelligence literature as analogous to that of biological neurons. The arrows leading to a node are like the axons leading to a neuron. Like the axons, they carry a signal to the neuron or node. The arrows leading away from a node are like the dendrites of a neuron, and they carry a signal away from a neuron or node. The neurons of a brain have far more complex interactions than those displayed in the diagram, however the developers of neural networks view neural networks as abstracting the most relevant features of neurons in the human brain.

Neural networks "learn" by adjusting the strength of the signal coming from nodes in the previous layer connecting to it. As the neural network better learns how to predict the target value from the input pattern, each of the connections between the input neurons and the hidden or intermediate neurons and between the intermediate neurons and the output neurons increases or decreases in strength. A function called a threshold or activation function modifies the signal coming into the hidden layer nodes. In the early days of neural networks, this function produced a value of 1 or 0, depending on whether the signal from the prior layer exceeded a threshold value. Thus, the node or neuron would only "fire" if the signal exceeded the threshold, a process thought to be similar to that of a neuron. It is now known that biological neurons are more complicated than previously believed. A simple all or none rule does not describe the behavior of biological neurons. Currently, activation functions are typically sigmoid in shape and can take on any value between 0 and 1 or between -1 and 1, depending on the particular function chosen. The modified signal is then output to the output layer nodes, which also apply activation functions. Thus, the information about the pattern being learned is encoded in the signals carried to and from the nodes. These signals map a relationship between the input nodes or the data and the output nodes or dependent variable.

<u>Example 1: Simple Example of Fitting a Nonlinear Function</u>

A simple example will be used to illustrate how neural networks perform nonlinear function approximations. This example will provide detail about the activation functions in the hidden and output layers to facilitate an understanding of how neural networks work.

In this example the true relationship between an input variable X and an output variable Y is exponential and is of the following form:

$$Y = e^{\frac{X}{2}} + \varepsilon$$

Where:

$\varepsilon \sim N(0,75)$

$X \sim N(12,.5)$

and N (μ, σ) is understood to denote the Normal probability distribution with parameters μ, the mean of the distribution and σ, the standard deviation of the distribution.

A sample of one hundred observations of X and Y was simulated. A scatterplot of the X and Y observations is shown in Figure 2. It is not clear from the scatterplot that the relationship between X and Y is nonlinear. The scatterplot in Figure 3 displays the "true" curve for Y as well as the random X and Y values.

Figure 2

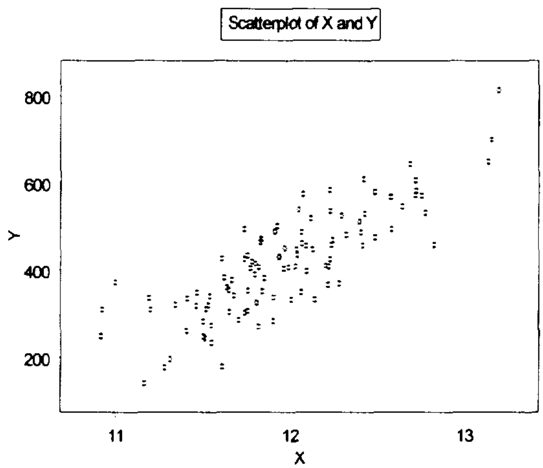
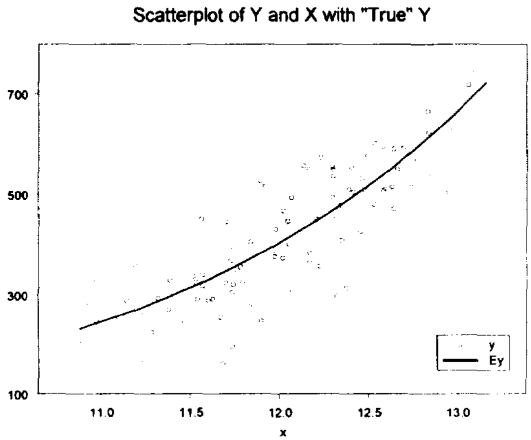
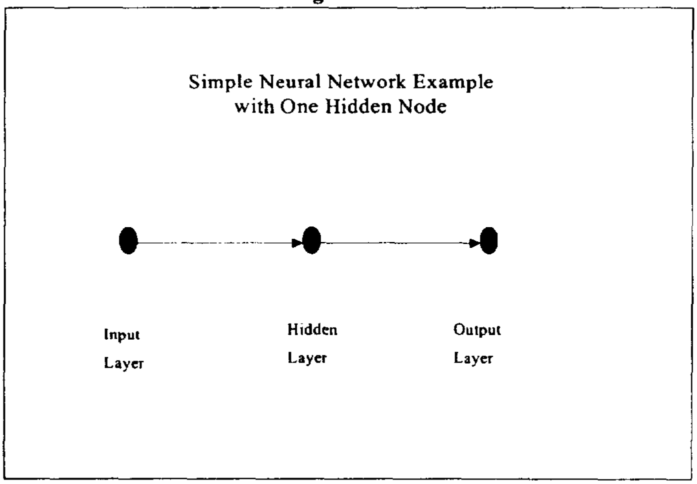


Figure 3



A simple neural network with one hidden layer was fit to the simulated data. In order to compare neural networks to classical models, a regression curve was also fit. The result of that fit will be discussed after the presentation of the neural network results. The structure of this neural network is shown in Figure 4.

Figure 4



As neural networks go, this is a relatively simple network with one input node. In biological neurons, electrochemical signals pass between neurons. In neural network analysis, the signal between neurons is simulated by software, which applies weights to the input nodes (data) and then applies an activation function to the weights.

Neuron signal of the biological neuron system → Node weights of neural networks

The weights are used to compute a linear sum of the independent variables. Let Y denote the weighted sum:

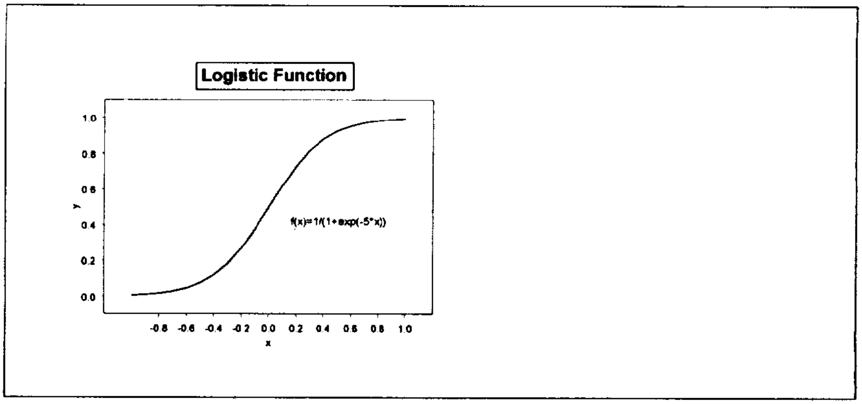
$$Y = w_0 + w_1 * X_1 + w_2 X_2 ... + w_n X_n$$

The activation function is applied to the weighted sum and is typically a sigmoid function. The most common of the sigmoid functions is the logistic function:

$$f(Y) = \frac{1}{1 + e^{-Y}}$$

The logistic function takes on values in the range 0 to 1. Figures 5 displays a typical logistic curve. This curve is centered at an X value of 0, (i.e., the constant w_0 is 0). Note that this function has an inflection point at an X value of 0 and f (x) value of .5, where it shifts from a convex to a concave curve. Also note that the slope is steepest at the inflection point where small changes in the value of X can produce large changes in the value of the function. The curve becomes relatively flat as X approaches both 1 and −1.

Figure 5



Another sigmoid function often used in neural networks is the hyperbolic tangent function which takes on values between –1 and 1:

$$f(Y) = \frac{e^{Y} - e^{-Y}}{e^{Y} + e^{-Y}}$$

In this paper, the logistic function will be used as the activation function. The Multilayer Perceptron is a multilayer feedforward neural network with a sigmoid activation function.

The logistic function is applied to the weighted input. In this example, there is only one input, therefore the activation function is:

$$h = f(X; w_{0,} w_1) = f(w_0 + w_1 X) = \frac{1}{1 + e^{-(w_0 + w_1 X)}}$$

This gives the value or activation level of the node in the hidden layer. Weights are then applied to the hidden node:

$$w_2 + w_3 h$$

The weights w_0 and w_2 are like the constants in a regression and the weights w_1 and w_3 are like the coefficients in a regression. An activation function is then applied to this "signal" coming from the hidden layer:

$$o = f(h; w_2, w_3) = \frac{1}{1 + e^{-(w_2 + w_3 h)}}$$

The output function o for this particular neural network with one input node and one hidden node can be represented as a double application of the logistic function:

$$f(f(X; w_0, w_1); w_3, w_4) = \frac{1}{1 + e^{-(w_3 + w_4 \frac{1}{1 + e^{-w_0 + w_1 X}})}}$$

It will be shown later in this paper that the use of sigmoid activation functions on the weighted input variables, along with the second application of a sigmoid function by the output node is what gives the MLP the ability to approximate nonlinear functions.

One other operation is applied to the data when fitting the curve: normalization. The dependent variable X is normalized. Normalization is used in statistics to minimize the impact of the scale of the independent variables on the fitted model. Thus, a variable with values ranging from 0 to 500,000 does not prevail over variables with values ranging from 0 to 10, merely because the former variable has a much larger scale.

Various software products will perform different normalization procedures. The software used to fit the networks in this paper normalizes the data to have values in the range 0 to 1. This is accomplished by subtracting a constant from each observation and dividing by a scale factor. It is common for the constant to equal the minimum observed value for X in the data and for the scale factor to equal the range of the observed values (the maximum minus the minimum). Note also that the output function takes on values between 0 and 1 while Y takes on values between $-\infty$ and $+\infty$ (although for all practical purposes, the probability of negative values for the data in this particular example is nil). In order to produce predicted values the output, o, must be renormalized by multiplying by a scale factor (the range of Y in our example) and adding a constant (the minimum observed Y in this example).

Fitting the Curve

The process of finding the best set of weights for the neural network is referred to as training or learning. The approach used by most commercial software to estimate the weights is backpropagation. Each time the network cycles through the training data, it produces a predicted value for the target variable. This value is compared to the actual value for the target variable and an error is computed for each observation. The errors are "fed back" through the network and new weights are computed to reduce the overall error. Despite the neural network terminology, the training process is actually a statistical optimization procedure. Typically, the procedure minimizes the sum of the squared residuals:

$$Min(\Sigma(Y - \hat{Y})^2)$$

Warner and Misra (Warner and Misra, 1996) point out that neural network analysis is in many ways like linear regression, which can be used to fit a curve to data. Regression coefficients are solved for by minimizing the squared deviations between actual observations on a target variable and the fitted value. In the case of linear regression, the curve is a straight line. Unlike linear regression, the relationship between the predicted and target variable in a neural network is nonlinear, therefore a closed form solution to the minimization problem does not exist. In order to minimize the loss function, a numerical technique such as gradient descent (which is similar to backpropagation) is used. Traditional statistical procedures such as nonlinear regression, or the solver in Excel use an approach similar to neural networks to estimate the parameters of nonlinear functions. A brief description of the procedure is as follows:

1. Initialize the neural network model using an initial set of weights (usually randomly chosen). Use the initialized model to compute a fitted value for an observation.
2. Use the difference between the fitted and actual value on the target variable to compute the error.

3. Change the weights by a small amount that will move them in the direction of a smaller error
    - This involves multiplying the error by the partial derivative of the function being minimized with respect to the weights. This is because the partial derivative gives the rate of change with respect to the weights. This is then multiplied by a factor representing the "learning rate" which controls how quickly the weights change. Since the function being approximated involves logistic functions of the weights of the output and hidden layers, multiple applications of the chain rule are needed. While the derivatives are a little messy to compute, it is straightforward to incorporate them into software for fitting neural networks.
4. Continue the process until no further significant reduction in the squared error can be obtained

Further details are beyond the scope of this paper. However, more detailed information is supplied by some authors (Warner and Misra, 1996, Smith, 1996). The manuals of a number of statistical packages (SAS Institute, 1988) provide an excellent introduction to several numerical methods used to fit nonlinear functions.

Fitting the Neural Network

For the more ambitious readers who wish to create their own program for fitting neural networks, Smith (Smith, 1996) provides an Appendix with computer code for constructing a backpropagation neural network. A chapter in the book computes the derivatives mentioned above, which are incorporated into the computer code.

However, the assumption for the purposes of this paper is that the overwhelming majority of readers will use a commercial software package when fitting neural networks. Many hours of development by advanced specialists underlie these tools. Appendix 1 discusses some of the software options available for doing neural network analysis.

The Fitted Curve:

The parameters fitted by the neural network are shown in Table 1.

Table 1

	w_0	w_1
Input Node to Hidden Node	-3.088	3.607
Hidden Node to Output Node	-1.592	5.281

To produce the fitted curve from these coefficients, the following procedure must be used:

1. Normalize each x_i by subtracting the minimum observed value [2] and dividing by the scale coefficient equal to the maximum observed X minus the minimum observed X. The normalized values will be denoted X*.
2. Determine the minimum observed value for Y and the scale coefficient for Y[3].
3. For each normalized observation x^*_i compute

$$h(x^*_i) = \frac{1}{1+e^{-(-3.088+3.607x^*_i)}}$$

4. For each h (x^*_i) compute

$$o(h(x^*_i)) = \frac{1}{1+e^{-(-1.592+5.281h(x^*_i))}}$$

5. Compute the estimated value for each y_i by multiplying the normalized value from the output layer in step 4 by the Y scale coefficient and adding the Y constant. This value is the neural network's predicted value for y_i.

Table 2 displays the calculation for the first 10 observations in the sample.

[2] 10.88 in this example. The scale parameter is 2.28

[3] In this example the Y minimum was 111.78 and the scale parameter was 697.04

Table 2

(1) Input X	(2) Pattern Y	(3) Normalized X	(4) Weighted X Input	(5) Logistic(Wt X)	(6) Weighted Node 2	(7) Logistic	(8) Rescaled Predicted
		((1)-10.88)/2.28	-3.088+3.607*(3)	1/(1+exp(-(4)))	-1.5916+5.2814*(5)	1/(1+exp(-(6)))	697.04*(7)+111.78
12.16	665.0	0.5613	-1.0634	0.2567	-0.2361	0.4413	419.4
11.72	344.6	0.3704	-1.7518	0.1478	-0.8109	0.3077	326.3
11.39	281.7	0.2225	-2.2854	0.0923	-1.1039	0.2490	285.3
12.02	423.9	0.4999	-1.2850	0.2167	-0.4471	0.3900	383.7
12.63	519.4	0.7679	-0.3184	0.4211	0.6323	0.6530	566.9
11.19	366.7	0.1359	-2.5978	0.0693	-1.2257	0.2269	270.0
13.06	697.2	0.9581	0.3678	0.5909	1.5294	0.8219	684.7
11.57	368.6	0.3011	-2.0020	0.1190	-0.9631	0.2763	304.3
11.73	423.6	0.3709	-1.7501	0.1480	-0.8098	0.3079	326.4
11.05	221.4	0.0763	-2.8128	0.0566	-1.2925	0.2154	261.9

Figure 6 provides a look under the hood at the neural network's fitted functions. The graph shows the output of the hidden layer node and the output layer node after application of the logistic function. The outputs of each node are an exponential-like curve, but the output node curve is displaced upwards by about .2 from the hidden node curve. Figure 7 displays the final result of the neural network fitting exercise: a graph of the fitted and "true" values of the dependent variables versus the input variable.

Figure 6

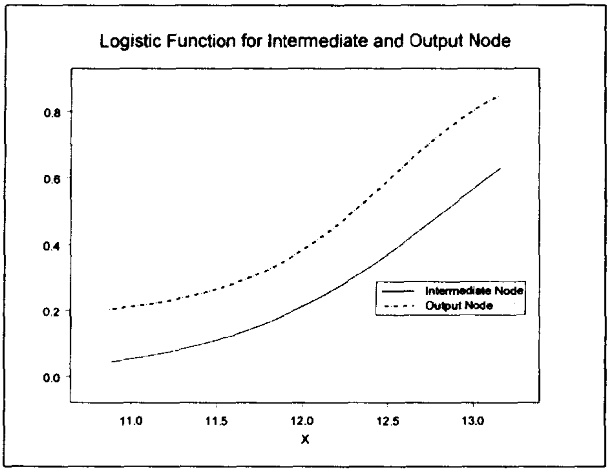
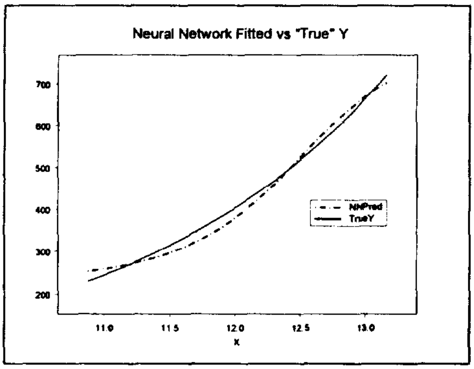


Figure 7



It is natural to compare this fitted value to that obtained from fitting a linear regression to the data. Two scenarios were used in fitting the linear regression. First, a simple straight line was fit, since the nonlinear nature of the relationship may not be apparent to the analyst. Since Y is an exponential function of X, the log transformation is a natural transformation for Y. However, because the error term in this relationship is additive, not multiplicative, applying the log transformation to Y produces a regression equation which is not strictly linear in both X and the error term:

$$Y = Ae^{B\frac{X}{2}} + \varepsilon \rightarrow \ln(Y) = \ln(Ae^{B\frac{X}{2}} + \varepsilon) \neq \ln(Y) = \ln(A) + B\frac{X}{2} + \varepsilon$$

Nonetheless, the log transformation should provide a better approximation to the true curve than fitting a straight line to the data. The regression using the log of Y as the dependent variable will be referred to as the exponential regression. It should be noted that the nonlinear relationship in this example could be fit using a nonlinear regression procedure which would address the concern about the log transform not producing a relationship which is linear in both X and ε. The purpose here, however, is to keep the exposition simple and use techniques that the reader is familiar with.

The table below presents the goodness of fit results for both regressions and the neural network. Most neural network software allows the user to hold out a portion of the sample for testing. This is because most modeling procedures fit the sample data better than they fit new observations presented to the model which were not in the sample. Both the neural network and the regression models were fit to the first 80 observations and then tested on the next 20. The mean of the squared errors for the sample and the test data is shown in Table 3

Table 3

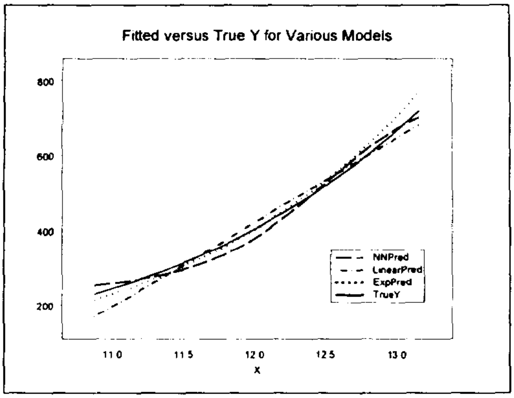
Method	Sample MSE	Test MSE
Linear Regression	4,766	8,795
Exponential Regression	4,422	7,537
Neural Network	4,928	6,930

As expected, all models fit the sample data better than they fit the test data. This table indicates that both of the regressions fit the sample data better than the neural network did, but the neural network fit the test data better than the regressions did.

The results of this simple example suggest that the exponential regression and the neural network with one hidden node are fairly similar in their predictive accuracy. In general, one would not use a neural network for this simple situation where there is only one predictor variable, and a simple transformation of one of the variables produces a curve which is a reasonably good approximation to the actual data. In addition, if the true function for the curve were known by the analyst, a nonlinear regression technique would probably provide the best fit to the data. However, in actual applications, the functional form of the relationship between the independent and dependent variable is often not known.

A graphical comparison of the fitted curves from the regressions, the neural network and the "true" values is shown in Figure 8.

Figure 8



The graph indicates that both the exponential regression and the neural network model provide a reasonably good fit to the data.

The logistic function revisited

The two parameters of the logistic function give it a lot of flexibility in approximating nonlinear curves. Figure 9 presents logistic curves for various values of the coefficient w_1. The coefficient controls the steepness of the curve and how quickly it approached its maximum and minimum values of 1 and -1. Coefficients with absolute values less than or equal to 1 produce curves which are straight lines. Figure 10 presents the effect of varying w_0 on logistic curves.

Figure 9

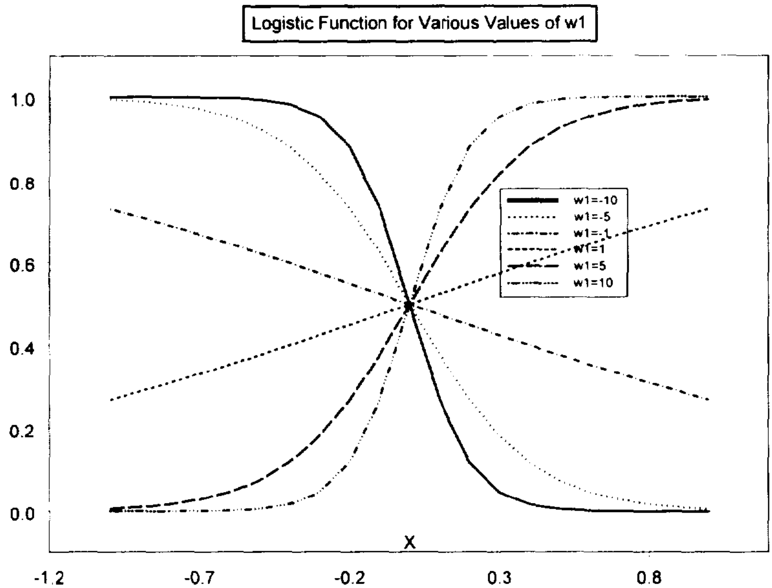
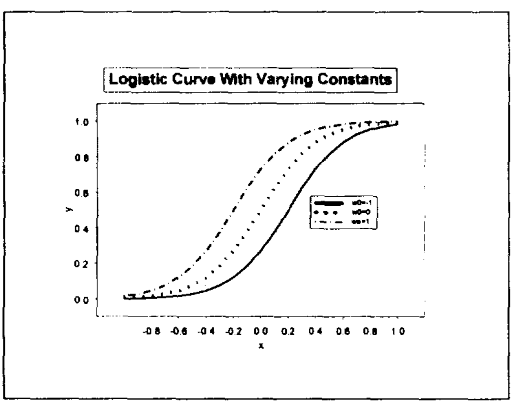
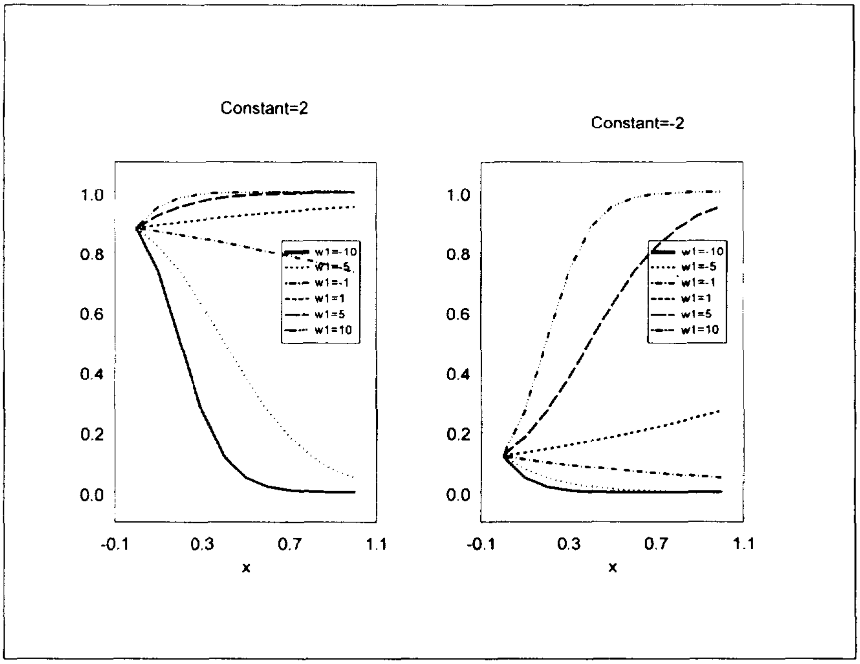


Figure 10



Varying the values of w_0 while holding w_1 constant shifts the curve right or left. A great variety of shapes can be obtained by varying the constant and coefficients of the logistic functions. A sample of some of the shapes is shown in Figure 11. Note that the X values on the graph are limited to the range of 0 to 1, since this is what the neural networks use. In the previous example the combination of shifting the curve and adjusting the steepness coefficient was used to define a curve that is exponential in shape in the region between 0 and 1.

Figure 11



Using Neural Networks to Fit a Complex Nonlinear Function:

To facilitate a clear introduction to neural networks and how they work, the first example in this paper was intentionally simple. The next example is a somewhat more complicated curve.

Example 2: A more complex curve
The function to be fit in this example is of the following form:

$$f(X) = \ln(X) + \sin(\frac{X}{675})$$

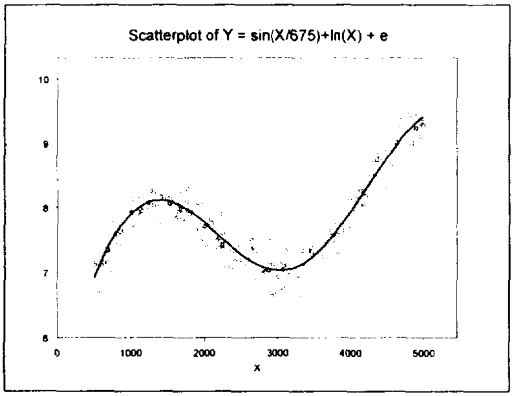
$$X \sim U(500{,}5000)$$

$$e \sim N(0,.2)$$

Note that U denotes the uniform distribution, and 500 and 5,000 are the lower and upper ends of the range of the distribution.

A scatterplot of 200 random values for Y along with the "true" curve are shown in Figure 12

Figure 12



This is a more complicated function to fit than the previous exponential function. It contains two "humps" where the curve changes direction. To illustrate how neural networks approximate functions, the data was fit using neural networks of different sizes. The results from fitting this curve using two hidden nodes will be described first. Table 4 displays the weights obtained from training for the two hidden nodes. W_0 denotes the constant and W_1 denotes the coefficient applied to the input data. The result of applying these weights to the input data and then applying the logistic function is the values for the hidden nodes.

Table 4		
	W_0	W_1
Node 1	-4.107	7.986
Node 2	6.549	-7.989

A plot of the logistic functions for the two intermediate nodes is shown below (Figure 13). The curve for Node 1 is S shaped, has values near 0 for low values of X and increases to values near 1 for high values of X. The curve for Node 2 is concave downward, has a value of 1 for low values of X and declines to about .2 at high values of X.

Figure 13

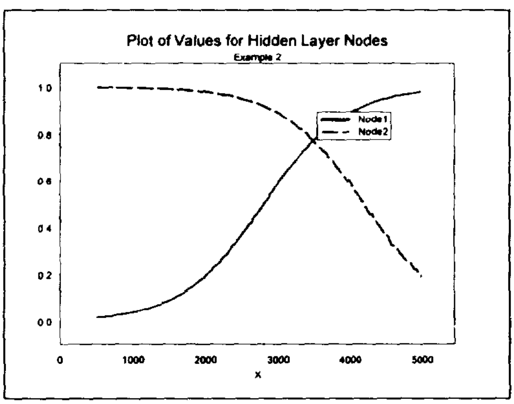


Table 5 presents the fitted weights connecting the hidden layer to the output layer:

	Table 5	
W_0	W_1	W_2
6.154	-3.0501	-6.427

Table 6 presents a sample of applying these weights to several selected observations from the training data to which the curve was fit. The table shows that the combination of the values for the two hidden node curves, weighted by the coefficients above produces a curve which is like a sine curve with an upward trend. At low values of X (about 500), the value of node 1 is low and node 2 is high. When these are weighted together, and the logistic function is applied, a moderately low value is produced. At values of X around 3,000, the values of both nodes 1 and 2 are relatively high. Since the coefficients of both nodes are negative, when they are weighted together, the value of the output function declines. At high values of X, the value of node 1 is high, but the value of node 2 is low. When the weight for node 1 is applied (-3.05) and is summed with the constant the value of the output node reduced by about 3. When the weight for node 2 (-6.43) is applied to the low output of node 2 (about .2) and the result is summed with the constant and the first node, the output node value is reduced by about 1 resulting in a weighted hidden node output of about 2. After the application of the logistic function the value of the output node is relatively high, i.e. near 1. Since the coefficient of node 1 has a lower absolute value, the overall result is a high value for the output function. Figure 14 presents a graph showing the values of the hidden nodes, the weighted hidden nodes (after the weights are applied to the hidden layer output but before the logistic function is applied) and the value of the output node (after the logistic function is applied to the weighted hidden node values). The figure shows how the application of the logistic function to the weighted output of the two hidden layer nodes produces a highly nonlinear curve.

Table 6						
Computation of Predicted Values for Selected Values of X						
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	((1)-508)/4994			6.15-3.05*(3)-6.43*(4)	1/(1+exp(-(5))	6.52+3.56*(6)
X	Normalized X	Output of Node 1	Output of Node 2	Weighted Hidden Node Output	Output Node Logistic Function	Predicted Y
508.48	0.00	0.016	0.999	-0.323	0.420	7.889
1,503.00	0.22	0.088	0.992	-0.498	0.378	7.752
3,013.40	0.56	0.596	0.890	-1.392	0.199	7.169
4,994.80	1.00	0.980	0.190	1.937	0.874	9.369

Figure 15 shows the fitted curve and the "true" curve for the two node neural network just described. One can conclude that the fitted curve, although producing a highly nonlinear curve, does a relatively poor job of fitting the curve for low values of X. It turns out that adding an additional hidden node significantly improves the fit of the curve.

Figure 14

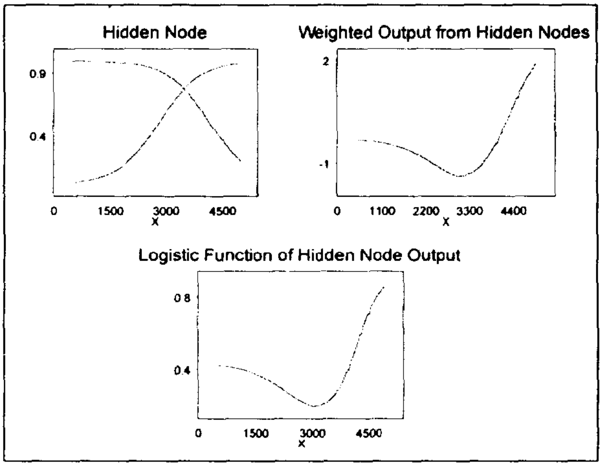


Figure 15

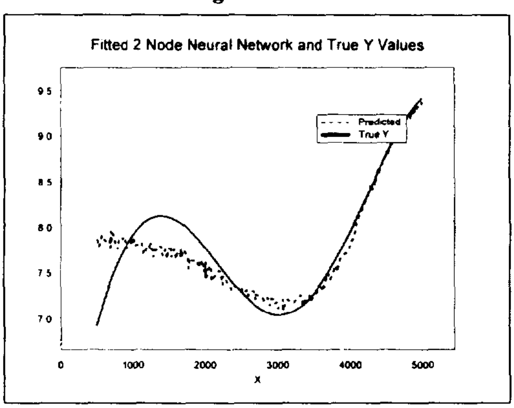


Table 7 displays the weights connecting the hidden node to the output node for the network with 3 hidden nodes. Various aspects of the hidden layer are displayed in Figure 16. In Figure 16, the graph labeled "Weighted Output of Hidden Node" displays the result of applying the Table 7 weights obtained from the training data to the output from the hidden nodes. The combination of weights, when applied to the three nodes produces a result which first increases, then decreases, then increases again. When the logistic function is applied to this output, the output is mapped into the range 0 to 1 and the curve appears to become a little steeper. The result is a curve that looks like a sine function with an increasing trend. Figure 17 displays the fitted curve, along with the "true" Y value.

	Table 7		
Weight 0	Weight 1	Weight 2	Weight 3
-4.2126	6.8466	-7.999	6.0722

Figure 16

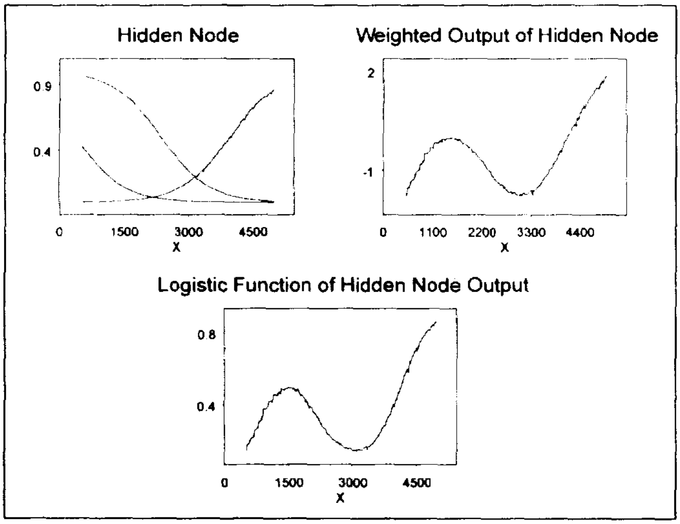
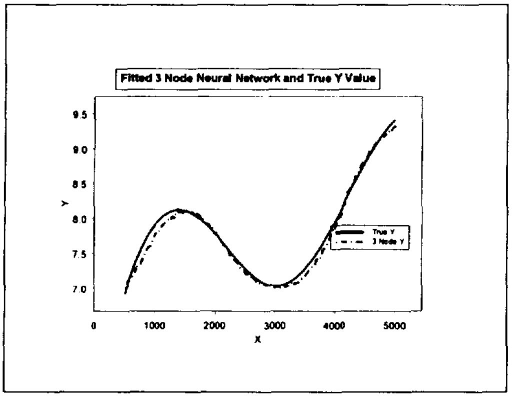


Figure 17



It is clear that the three node neural network provides a considerably better fit than the two node network. One of the features of neural networks which affects the quality of the fit and which the user must often experiment with is the number of hidden nodes. If too many hidden nodes are used, it is possible that the model will be overparameterized. However, an insufficient number of nodes could be responsible for a poor approximation of the function.

This particular example has been used to illustrate an important feature of neural networks: the multilayer perceptron neural network with one hidden layer is a universal function approximator. Theoretically, with a sufficient number of nodes in the hidden layer, any nonlinear function can be approximated. In an actual application on data containing random noise as well as a pattern, it can sometimes be difficult to accurately approximate a curve no matter how many hidden nodes there are. This is a limitation that neural networks share with classical statistical procedures.

Neural networks are only one approach to approximating nonlinear functions. A number of other procedures can also be used for function approximation. A conventional statistical approach to fitting a curve to a nonlinear function when the form of the function is unknown is to fit a polynomial regression:

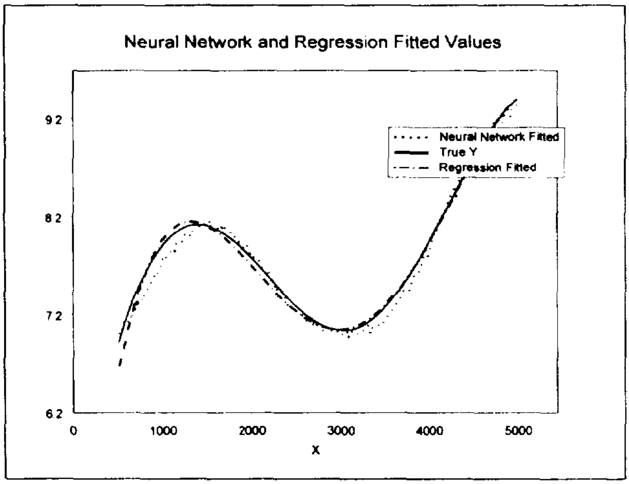
$$Y = a + b_1 X + b_2 X^2 ... + b_n X^n$$

Using polynomial regression, the function is approximated with an n^{th} degree polynomial. Higher order polynomials are used to approximate more complex functions. In many situations polynomial approximation provides a good fit to the data. Another advanced method for approximating nonlinear functions is to fit regression splines. Regression splines fit piecewise polynomials to the data. The fitted polynomials are constrained to have second derivatives at each breakpoint; hence a smooth curve is produced. Regression splines are an example of contemporary data mining tools and will not be discussed further in this paper. Another function approximator that actuaries have some familiarty with is the Fourier transform which uses combinations of sine and cosine functions to approximate curves. Among actuaries, their use has been primarily to approximate aggregate loss distributions. Heckman and Meyers (Heckman and Meyers, 1983) popularized this application.

In this paper, since neural networks are being compared to classical statistical procedures, the use of polynomial regression to approximate the curve will be illustrated. Figure 18 shows the result of fitting a 4^{th} degree polynomial curve to the data from Example 2. This is the polynomial curve which produced the best fit to the data. It can be concluded from Figure 18 that the polynomial curve produces a good fit to the data. This is not surprising given that using a Taylor series approximation both the sine function and log function can be approximated relatively accurately by a series of polynomials.

Figure 18 allows the comparison of both the Neural Network and Regression fitted values. It can be seen from this graph that both the neural network and regression provide a reasonable fit to the curve.

Figure 18



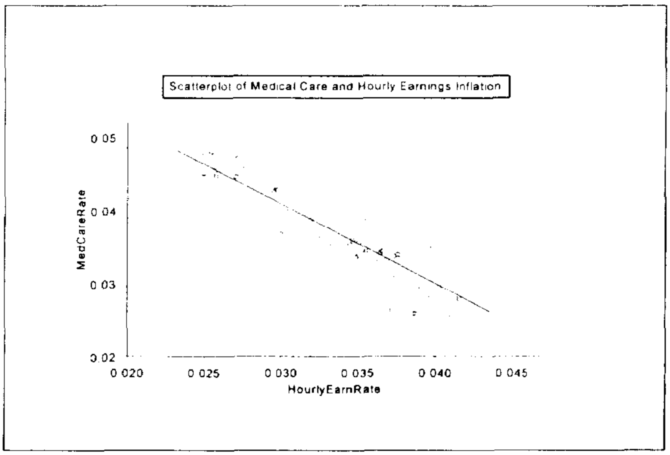
While these two models appear to have similar fits to the simulated nonlinear data, the regression slightly outperformed the neural network in goodness of fit tests. The r^2 for the regression was higher for both training (.993 versus .986) and test (.98 versus .94) data.

Correlated Variables and Dimension Reduction

The previous sections discussed how neural networks approximate functions of a variety of shapes and the role the hidden layer plays in the approximation. Another task performed by the hidden layer of neural networks will be discussed in this section: dimension reduction.

Data used for financial analysis in insurance often contains variables that are correlated. An example would be the age of a worker and the worker's average weekly wage, as older workers tend to earn more. Education is another variable which is likely to be correlated with the worker's income. All of these variables will probably influence Workers Compensation indemnity payments. It could be difficult to isolate the effect of the individual variables because of the correlation between the variables. Another example is the economic factors that drive insurance inflation, such as inflation in wages and inflation in the medical care. For instance, analysis of monthly Bureau of Labor Statistics data for hourly wages and the medical care component of the CPI from January of 1994 through May of 2000 suggest these two time series have a (negative) correlation of about .9 (See Figure 19). Other measures of economic inflation can be expected to show similarly high correlations.

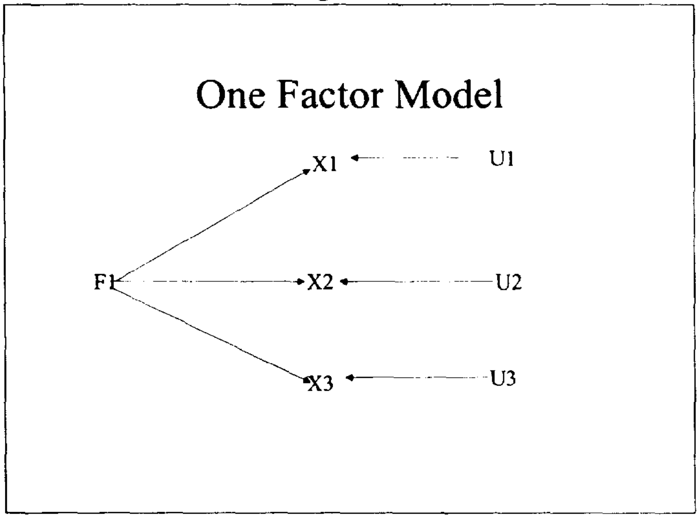
Figure 19



Suppose one wanted to combine all the demographic factors related to income level or all the economic factors driving insurance inflation into a single index in order to create a simpler model which captured most of the predictive ability of the individual data series. Reducing many factors to one is referred to as dimension reduction. In classical statistics, two similar techniques for performing dimension reduction are Factor Analysis and Principal Components Analysis. Both of these techniques take a number of correlated variables and reduce them to fewer variables which retain most of the explanatory power of the original variables.

The assumptions underlying Factor Analysis will be covered first. Assume the values on three observed variables are all "caused" by a single factor plus a factor unique to each variable. Also assume that the relationships between the factors and the variables are linear. Such a relationship is diagrammed in Figure 20, where F1 denotes the common factor, U1, U2 and U3 the unique factors and X1, X2 and X3 the variables. The causal factor F1 is not observed. Only the variables X1, X2 and X3 are observed. Each of the unique factors is independent of the other unique factors, thus any observed correlations between the variables is strictly a result of their relation to the causal factor F1.

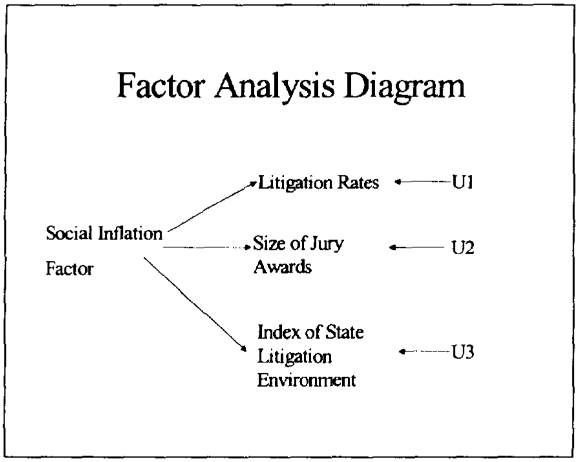
Figure 20



For instance, assume an unobserved factor, social inflation, is one of the drivers of increases in claims costs. This factor reflects the sentiments of large segments of the population towards defendants in civil litigation and towards insurance companies as intermediaries in liability claims. Although it cannot be observed or measured, some of its effects can be observed. Examples are the change over time in the percentage of claims being litigated, increases in jury awards and perhaps an index of the litigation environment in each state created by a team of lawyers and claims adjusters. In the social sciences it is common to use Factor Analysis to measure social and psychological concepts that cannot be directly observed but which can influence the outcomes of variables that can be directly observed. Sometimes the observed variables are indices or scales obtained from survey questions.

The social inflation scenario might be diagrammed as follows:

Figure 21



In scenarios such as this one, values for the observed variables might be used to obtain estimates for the unobserved factor. One feature of the data that is used to estimate the factor is the correlations between the observed variables. If there is a strong relationship between the factor and the variables, the variables will be highly correlated. If the relationship between the factor and only two of the variables is strong, but the relationship with the third variable is weak, then only the two variables will have a high correlation. The highly correlated variables will be more important in estimating the unobserved factor. A result of Factor Analysis is an estimate of the factor (F1) for each of the observations. The F1 obtained for each observation is a linear combination of the values for the three variable for the observation. Since the values for the variables will differ from record to record, so will the values for the estimated factor.

Principal Components Analysis is in many ways similar to Factor Analysis. It assumes that a set of variables can be described by a smaller set of factors which are linear combinations of the variables. The correlation matrix for the variables is used to estimate these factors. However, Principal Components Analysis makes no assumption about a causal relationship between the factors and the variables. It simply tries to find the factors or components which seem to explain most of the variance in the data. Thus both Factor Analysis and Principal Components Analysis produce a result of the form:

$$\hat{I} = w_1 X_1 + w_2 X_2 ... + w_n X_n$$

where

$\hat{I}$ is an estimate of the index or factor being constructed
X_1 ..X_n are the observed variables used to construct the index
w_1..w_n are the weights applied to the variables

An example of creating an index from observed variables is combining observations related to litigiousness and the legal environment to produce a social inflation index. Another example is combining economic inflationary variables to construct an economic inflation index for a line of business.[4] Factor analysis or Principal Components Analysis can be used to do this. Sometimes the values observed on variables are the result of or "caused" by more than one underlying factor. The Factor Analysis and Principal Components approach can be generalized to find multiple factors or indices, when the observed variables are the result of more than one unobserved factor.

One can then use these indices in further analyses and discard the original variables. Using this approach, the analyst achieves a reduction in the number of variables used to model the data and can construct a more parsimonious model.

Factor Analysis[5] is an example of a more general class of models known as Latent Variable Models. For instance, observed values on categorical variables may also be the result of unobserved factors. It would be difficult to use Factor Analysis to estimate the underlying factors because it requires data from continuous variables, thus an alternative procedure is required. While a discussion of such procedures is beyond the scope of this paper, the procedures do exist.

It is informative to examine the similarities between Factor Analysis and Principal Components Analysis and neural networks. Figure 22 diagrams the relationship between input variables, a single unobserved factor and the dependent variable. In the scenario diagrammed, the input variables are used to derive a single predictive index (F1) and the index is used to predict the dependent variable. Figure 23 diagrams the neural network being applied to the same data. Instead of a factor or index, the neural network has a hidden layer with a single node. The Factor Analysis index is a weighted linear combination of the input variables, while in the typical MLP neural network, the hidden layer is a weighted nonlinear combination of the input variables. The dependent variable is a linear function of the Factor in the case of Factor Analysis and Principal Components Analysis and (possibly) a non linear function of the hidden layer in the case of the MLP. Thus, both procedures can be viewed as performing dimension reduction. In the case of

[4] In fact Masterson created such indices for the Property and Casualty lines in the 1960s.
[5] Principal Components, because it does not have an underlying causal factor is not a latent variable model.

neural networks, the hidden layer performs the dimension reduction. Since it is performed using nonlinear functions, it can be applied where nonlinear relationships exist.

Example 3: Dimension reduction
Both Factor Analysis and neural networks will be fit to data where the underlying relationship between a set of independent variables and a dependent variable is driven by an underlying unobserved factor. An underlying causal factor, *Factor*1, is generated from a normal distribution:

$$Factor1 \sim N(1.05,.025)$$

On average this factor produces a 5% inflation rate. To make this example concrete *Factor*1 will represent the economic factor driving the inflationary results in a line of business, say Workers Compensation. *Factor*1 drives the observed values on three simulated economic variables, Wage Inflation, Medical Inflation and Benefit Level Inflation. Although unrealistic, in order to keep this example simple it was assumed that no factor other than the economic factor contributes to the value of these variables and the relationship of the factors to the variables is approximately linear.

Figure 22

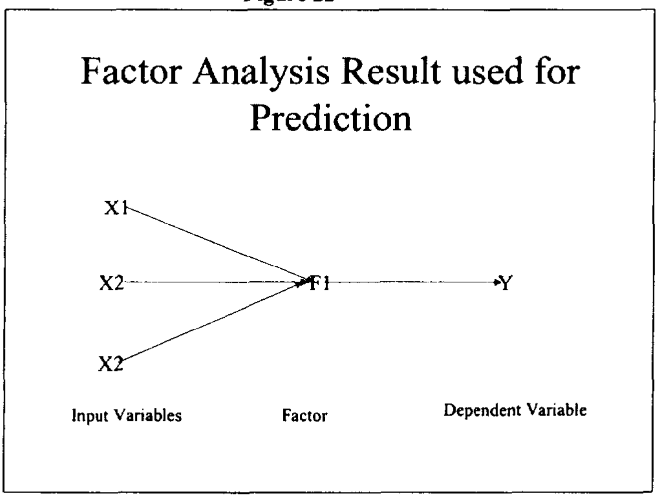
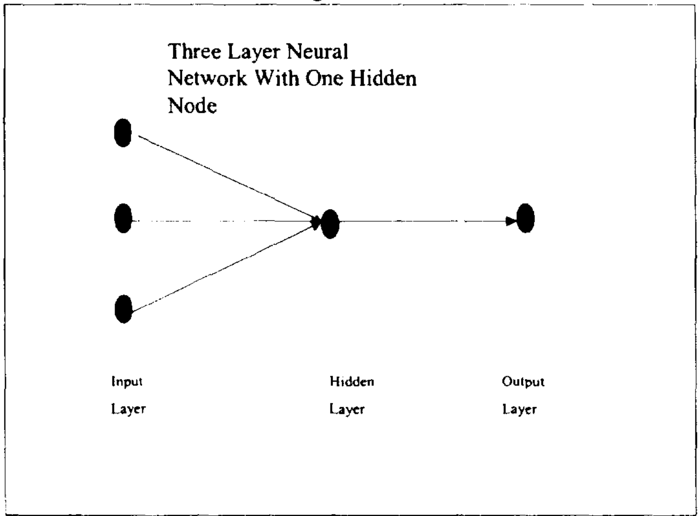


Figure 23



Also, to keep the example simple it was assumed that one economic factor drives Workers Compensation results. A more realistic scenario would separately model the indemnity and medical components of Workers Compensation claim severity. The economic variables are modeled as follows[6]:

$$\ln(WageInflation) = .7 * \ln(Factor1) + e$$
$$e \sim N(0,.005)$$

$$\ln(MedicalInflation) = 1.3 * \ln(Factor1) + e$$
$$e \sim N(0,.01)$$

$$\ln(Benefit\_level\_trend) = .5 * \ln(Factor1) + e$$
$$e \sim N(0,.005)$$

Two hundred fifty records of the unobserved economic inflation factor and observed inflation variables were simulated. Each record represented one of 50 states for one of 5 years. Thus, in the simulation, inflation varied by state and by year. The annual inflation rate variables were converted into cumulative inflationary measures (or indices). For each state, the cumulative product of that year's factor and that year's observed inflation

[6] Note that the according to Taylor's theorem the natural log of a variable whose value is close to one is approximately equal to 1 minus the variable's value, i.e., $\ln(1+x) \approx x$. Thus, the economic variables are, to a close approximation, linear functions of the factor.

measures (the random observed variables) were computed. For example the cumulative unobserved economic factor is computed as:

$$Cumfactor1_t = \prod_{k=1}^{t} Factor1_k$$

A base severity, intended to represent the average severity over all claims for the line of business for each state for each of the 5 years was generated from a lognormal distribution. [7] To incorporate inflation into the simulation, the severity for a given state for a given year was computed as the product of the simulated base severity and the cumulative value for the simulated (unobserved) inflation factor for its state. Thus, in this simplified scenario, only one factor, an economic factor is responsible for the variation over time and between states in average severity. The parameters for these variables were selected to make a solution using Factor Analysis or Principal Components Analysis straightforward and are not based on an analysis of real insurance data. This data therefore had significantly less variance than would be observed in actual insurance data.

Note that the correlations between the variables is very high. All correlations between the variables are at least .9. This means that the problem of multicollineariy exists in this data set. That is, each variable is nearly identical to the others, adjusting for a constant multiplier, so typical regression procedures have difficulty estimating the parameters of the relationship between the independent variables and severity. Dimension reduction methods such as Factor Analysis and Principal Components Analysis address this problem by reducing the three inflation variables to one, the estimated factor or index.

Factor Analysis was performed on variables that were standardized. Most Factor Analysis software standardizes the variables used in the analysis by subtracting the mean and dividing by the standard deviation of each series. The coefficients linking the variables to the factor are called loadings. That is:

$X1 = b_1$ Factor1
$X2 = b_2$ Factor1
$X3 = b_3$ Factor1

Where X1, X2 and X3 are the three observed variables, Factor1 is the single underlying factor and b_1, b_2 and b_3 are the loadings.

In the case of Factor Analysis the loadings are the coefficients linking a standardized factor to the standardized dependent variables, not the variables in their original scale. Also, when there is only one factor, the loadings also represent the estimated correlations between the factor and each variable. The loadings produced by the Factor Analysis procedure are shown in Table 8.

---

[7] This distribution will have an average of 5,000 the first year (after application of the inflationary factor for year 1). Also $\ln(Severity) \sim N(8.47,.05)$

Table 8		
Variable	Loading	Weights
Wage Inflation Index	.985	.395
Medical Inflation Index	.988	.498
Benefit Level Inflation Index	.947	.113

Table 8 indicates that all the variables have a high loading on the factor, and thus all are likely to be important in the estimation of an economic index. An index value was estimated for each record using a weighted sum of the three economic variables. The weights used by the Factor Analysis procedure to compute the index are shown in Table 8. Note that these weights (within rounding error) sum to 1. The new index was then used as a dependent variable to predict each state's severity for each year. The regression model was of the form:

Index =.395 (Wage Inflation)+.498(Medical Inflation)+.113(Benefit Level Inflation)

$$Severity = a + b * Index + e$$

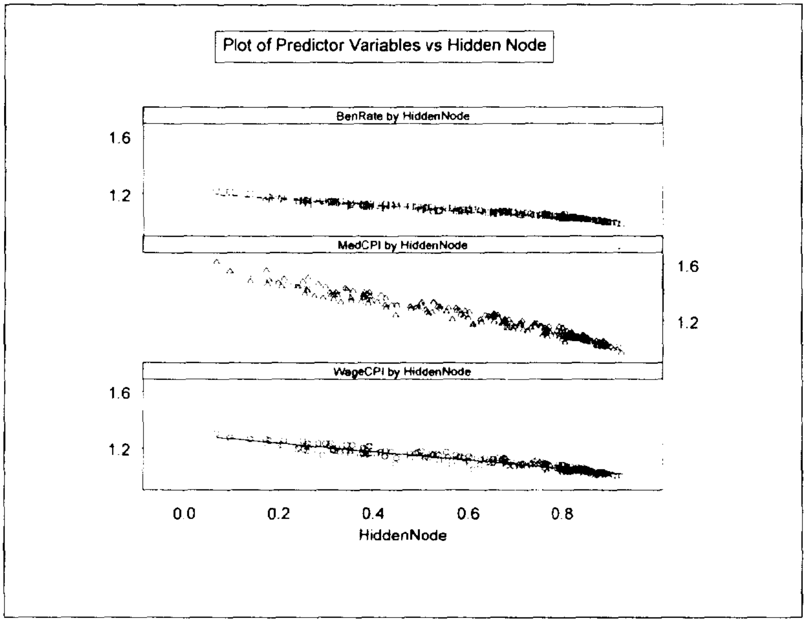
where

Severity is the simulated severity
Index is the estimated inflation Index from the Factor Analysis procedure
e is a random error term

The results of the regression will be discussed below where they are compared to those of the neural network.

The simple neural network diagramed in Figure 23 with three inputs and one hidden node was used to predict a severity for each state and year. Figure 24 displays the relationship between the output of the hidden layer and each of the predictor variables. The hidden node has a linear relationship with each of the independent variables, but is negatively correlated with each of the variables. The relationship between the neural network predicted value and the independent variables is shown in Figure 25. This relationship is linear and positively sloped. The relationship between the unobserved inflation factor driving the observed variables and the predicted values is shown in Figure 26. This relationship is positively sloped and nearly linear. Thus, the neural network has produced a curve which is approximately the same form as the "true" underlying relationship.

Figure 24



Plot of Predictor Variables vs Hidden Node
BenRate by HiddenNode
1.6
1.2
MedCPI by HiddenNode
1.6
1.2
WageCPI by HiddenNode
1.6
1.2
0.0
0.2
0.4
0.6
0.8
HiddenNode

Figure 25

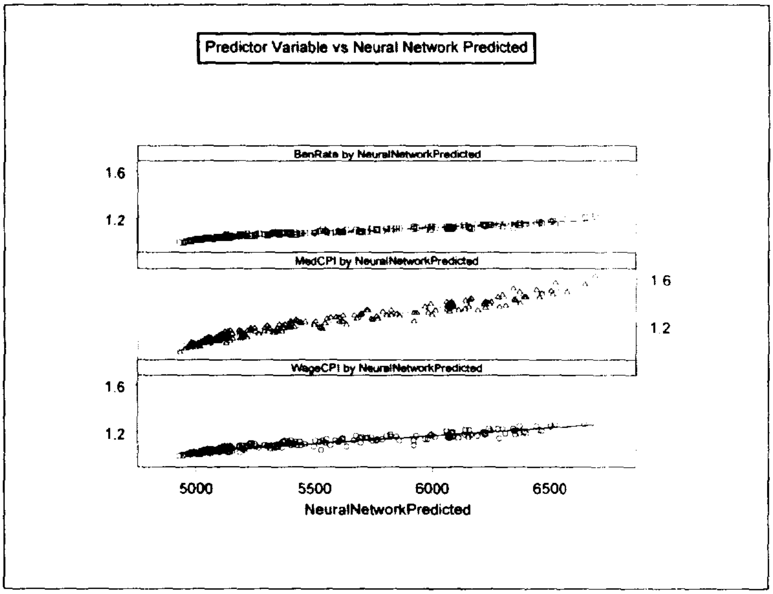
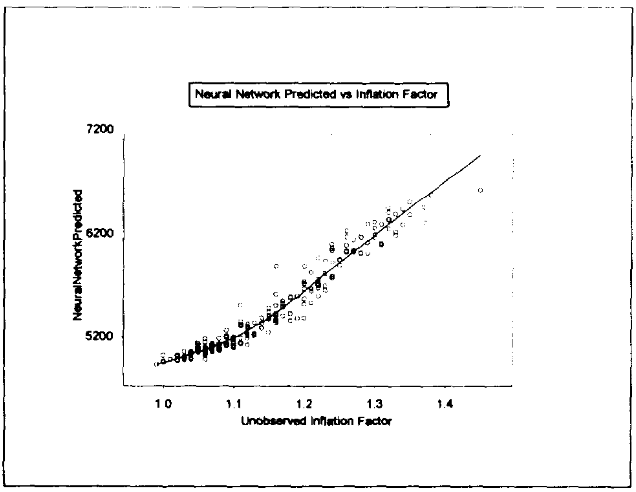


Figure 26



Interpreting the Neural Network Model

With Factor Analysis, a tool is provided for assessing the influence of a variable on a Factor and therefore on the final predicted value. The tool is the factor loadings which show the strength of the relationship between the observed variable and the underlying factor. The loadings can be used to rank each variable's importance. In addition, the weights used to construct the index[8] reveal the relationship between the independent variables and the predicted value (in this case the predicted value for severity).

Because of the more complicated functions involved in neural network analysis, interpretation of the variables is more challenging. One approach (Potts, 1999) is to examine the weight connecting the input variables to the hidden layer. Those which are closest to zero are least important. A variable is deemed unimportant only if all of these connections are near zero. Table 9 displays the values for the weights connecting the input layer to the hidden layer. Using this procedure, no variable in this example would be deemed "unimportant". This procedure is typically used to eliminate variables from a model, not to quantify their impact on the outcome. While it was observed above that application of these weights resulted in a network that has an approximate linear relationship with the predictor variables, the weights are relatively uninformative for determining the influence of the variables on the fitted values.

Table 9: Factor Example Parameters

W_0	W_1	W_2	W_3
2.549	-2.802	-3.010	0.662

Another approach to assessing the predictor variables' importance is to compute a sensitivity for each variable (Potts, 1999). The sensitivity is a measure of how much the predicted value's error increases when the variables are excluded from the model one at a time. However, instead of actually excluding variables, they are fixed at a constant value. The sensitivity is computed as follows:

1. Hold one of the variables constant; say at its mean or median value.
2. Apply the fitted neural network to the data with the selected variable held constant.
3. Compute the squared errors for each observation produced by these modified fitted values.
4. Compute the average of the squared errors and compare it to the average squared error of the full model.
5. Repeat this procedure for each variable used by the neural network. The sensitivity is the percentage reduction in the error of the full model, compared to the model excluding the variable in question.
6. If desired, the variables can be ranked based on their sensitivities.

[8] This would be computed as the product of each variable's weight on the factor times the coefficient of the factor in a linear regression on the dependent variable (.85 in this example).

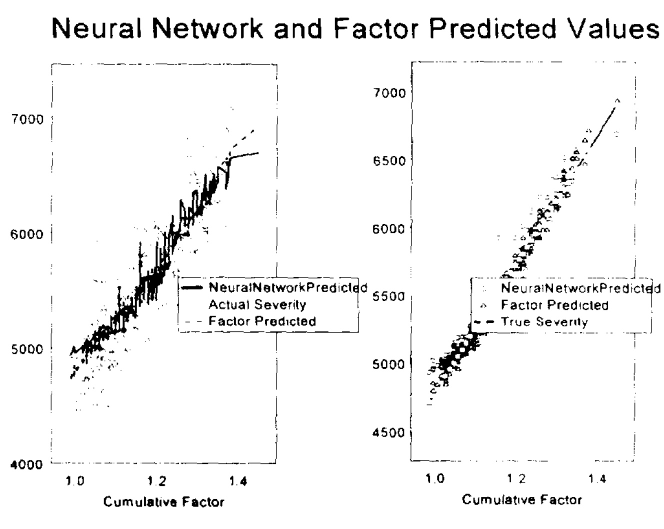
Since the same set of parameters is used to compute the sensitivities, this procedure does not require the user to refit the model each time a variable's importance is being evaluated. The following table presents the sensitivities of the neural network model fitted to the factor data.

Table 10	
Sensitivities of Variables in Factor Example	
Benefit Level	23.6%
Medical Inflation	33.1%
Wage Inflation	6.0%

According to the sensitivities, Medical Inflation is the most important variable followed by Benefit Level and Wage Inflation is the least important. This contrasts with the importance rankings of Benefit Level and Wage Inflation in the Factor Analysis, where Wage Inflation was a more important variable than Benefit Level. Note that these are the sensitivities for the particular neural network fit. A different initial starting point for the network or a different number of hidden nodes could result in a model with different sensitivities.

Figure 27 shows the actual and fitted values for the neural network and Factor Analysis predicted models. This figure displays the fitted values compared to actual randomly generated severities (on the left) and to "true" expected severities on the right. The x-axis of the graph is the "true" cumulative inflation factor, as the severities are a linear

Figure 27



function of the factor. However, it should be noted that when working with real data, information on an unobserved variable would not be available.

The predicted neural network values appear to be more jagged than the Factor Analysis predicted values. This jaggedness may reflect a weakness of neural networks: over fitting. Sometimes neural networks do not generalize as well as classical linear models, and fit some of the noise or randomness in the data rather than the actual patterns. Looking at the graph on the right showing both predicted values as well as the "true" value, the Factor Analysis model appears to be a better fit as it has less dispersion around the "true" value. Although the neural network fit an approximately linear model to the data, the Factor Analysis model performed better on the data used in this example. The Factor Analysis model explained 73% of the variance in the training data compared to 71% explained by the neural network model and 45% of the variance in the test data compared to 32% for the neural network. Since the relationships between the independent and dependent variables in this example are approximately linear, this is another instance of a situation where a classical linear model would be preferred over a more complicated neural network procedure.

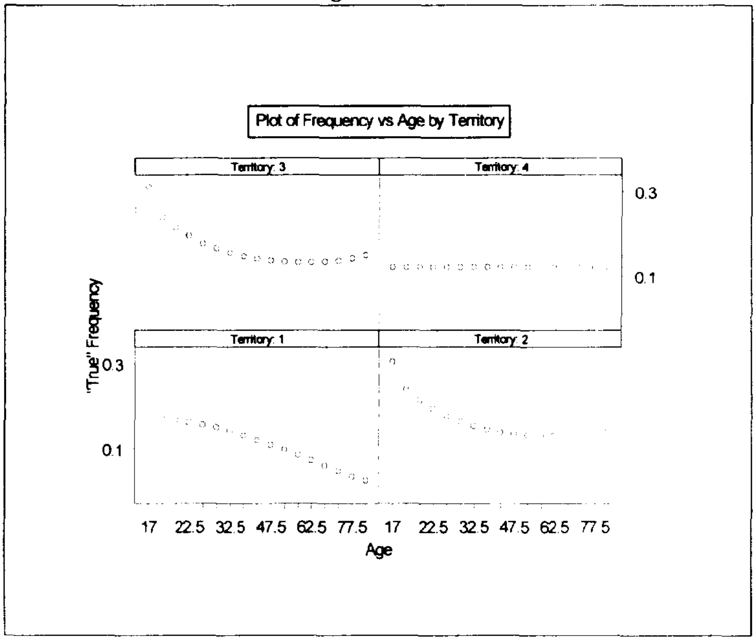
Interactions

Another common feature of data which complicates the statistical analysis is interactions. An interaction occurs when the impact of two variables is more or less than the sum of their independent impacts. For instance, in private passenger automobile insurance, the driver's age may interact with territory in predicting accident frequencies. When this happens, youthful drivers have a higher accident frequency in some territories than that given by multiplying the age and territory relativities. In other territories it is lower. An example of this is illustrated in Figure 28, which shows hypothetical curves[9] of expected or "true"(not actual) accident frequencies by age for each of four territories.

The graph makes it evident that when interactions are present, the slope of the curve relating the dependent variable (accident frequency) to an independent variable varies based on the values of a third variable (territory). It can be seen from the figure that younger drivers have a higher frequency of accidents in territories 2 and 3 than in territories 1 and 4. It can also be seen that in territory 4, accident frequency is not related to age and the shape and slope of the curve is significantly different in Territory 1 compared to territories 2 and 3.

[9] The curves are based on simulated data. However data from the Baxter (Venebles and Ripley) automobile claims database was used to develop parameters for the simulation.

Figure 28


Plot of Frequency vs Age by Territory
Territory: 3
Territory: 4
Territory: 1
Territory: 2
"True" Frequency
0.3
0.1
0.3
0.1
17 22.5 32.5 47.5 62.5 77.5 17 22.5 32.5 47.5 62.5 77.5
Age

As a result of interactions, the true expected frequency cannot be accurately estimated by the simple product of the territory relativity times the age relativity. The interaction of the two terms, age and territory, must be taken into account. In linear regression, interactions are estimated by adding an interaction term to the regression. For a regression in which the classification relativities are additive:

$Y_{ta} = B_0 + (B_t * \text{Territory}) + (B_a * \text{Age}) + (B_{at} * \text{Territory} * \text{Age})$

Where:
Y_{ta} = is either a pure premium or loss ratio for territory t and age a
B_0 = the regression constant
B_t, B_a and B_{at} are coefficients of the Territory, Age and the Age, Territory interaction

It is assumed in the regression model above that Territory enters the regression as a categorical variable. That is, if there are N territories, N-1 dummy variables are created which take on values of either 1 or 0, denoting whether an observation is or is not from each of the territories. One territory is selected as the base territory, and a dummy variable is not created for it. The value for the coefficient B_0 contains the estimate of the impact of the base territory on the dependent variable. More complete notation for the regression with the dummy variables is:

$Y_{ta} = B_0 + B_{t1}*T1 + B_{t2}*T2 + B_{t3} * T3 + B_a*Age + B_{at1}* T1*Age + B_{at2}* T2*Age + B_{at3}* T3*Age$

where T1, T2 and T3 are the dummy variables with values of either 1 or 0 described above and B_{t1} - B_{t3} are the coefficients of the dummy variables and B_{at1} - B_{at3}* are coefficients of the age and territory interaction terms. Note that most major statistical packages handle the details of converting categorical variables to a series of dummy variables.

The interaction term represents the product of the territory dummy variables and age. Using interaction terms allows the slope of the fitted line to vary by territory. A similar formula to that above applies if the class relativities are multiplicative rather than additive; however, the regression would be modeled on a log scale:

$\ln(Y_{ta}) = B^*_0 + (B^*_t * \text{Territory}) + (B^*_a * \text{Age}) + (B^*_{at} * \text{Territory} * \text{Age})$

where
B^*_0, B^*_t, B^*_a and B^*_{at} are the log scale constant and coefficients of the Territory, Age and Age, Territory interaction.

<u>Example 3: Interactions</u>
To illustrate the application of both neural networks and regression techniques to data where interactions are present 5,000 records were randomly generated. Each record represents a policyholder. Each policyholder has an underlying claim propensity dependent on his/her simulated age and territory, including interactions between these

two variables. The underlying claim propensity for each age and territory combination was that depicted above in Figure 28. For instance, in territory 4 the claim frequency is a flat .12. In the other territories the claim frequency is described by a curve. The claim propensity served as the Poisson parameter for claims following the Poisson distribution:

$$P(X = x; \lambda_{ij}) = \frac{\lambda_{ij}^{\ x}}{x!} e^{\lambda_{ij}}$$

Here λ_{ij} is the claim propensity or expected claim frequency for each age, territory combination. The claim propensity parameters were used to generate claims from the Poisson distribution for each of the 5,000 policyholders.[10]

Models for count data

The claims prediction procedures described in this section apply models to data with discrete rather than continuous outcomes. A policy can be viewed as having two possible outcomes: a claim occurs or a claim does not occur. We can assign the value 1 to observations with a claim and 0 to observations without a claim. The probability the policy will have a value of 1 lies in the range 0 to 1. When modeling such variables, it is useful to use a model where the possible values for the dependent variable lie in this range. One such modeling technique is logistic regression. The target variable is the probability that a given policyholder will have a claim, and this probability is denoted $p(\mathbf{x})$. The model relating $p(x)$ to the a vector of independent variables **x** is:

$$\ln(\frac{p}{1-p}; \mathbf{x}) = B_0 + B_1 X_1 + \ldots + B_n X_n$$

where the quantity $\ln(p(\mathbf{x})/(1-p(\mathbf{x})))$ is known as the logit function.

In general, specialized software is required to fit a logistic regression to data, since the logit function is not defined on individual observations when these observations can take on only the values 0 or 1. The modeling techniques work from the likelihood functions, where the likelihood function for a single observation is:

$$l(\mathbf{x}_i) = p(x_i)^{y_i}(1 - p(x_i)^{1-y_i})$$

$$p(x_i) = \frac{1}{1 + e^{-(B_0 + B_1 x_{i1} \ldots B_n x_{in})}}$$

Where $x_{i1} \ldots x_{in}$ are the independent variables for observation i, y_i is the response (either 0 or 1) and $B_1..B_n$ are the coefficients of the independent variables in the logistic regression. This logistic function is similar to the activation function used by neural networks. However, the use of the logistic function in logistic regression is very different from its use in neural networks. In logistic regression, a transform, the logit transform, is

[10] The overall distribution of drivers by age used in the simulation was based on fitting a curve to information from the US Department of Transportation web site.

applied to a target variable modeling it directly as a function of predictor variables. After parameters have been fit, the function can be inverted to produce fitted frequencies. The logistic functions in neural networks have no such straightforward interpretation. Numerical techniques are required to fit logistic regression when the maximum likelihood technique is used. Hosmer and Lemshow (Hosmer and Lemshow, 1989) provide a clear but detailed description of the maximum likelihood method for fitting logistic regression. Despite the more complicated methods required for fitting the model, in many other ways, logistic regression acts like ordinary least squares regression, albeit, one where the response variable is binary. In particular, the logit of the response variable is a linear function of the independent variables. In addition interaction terms, polynomial terms and transforms of the independent variables can be used in the model.

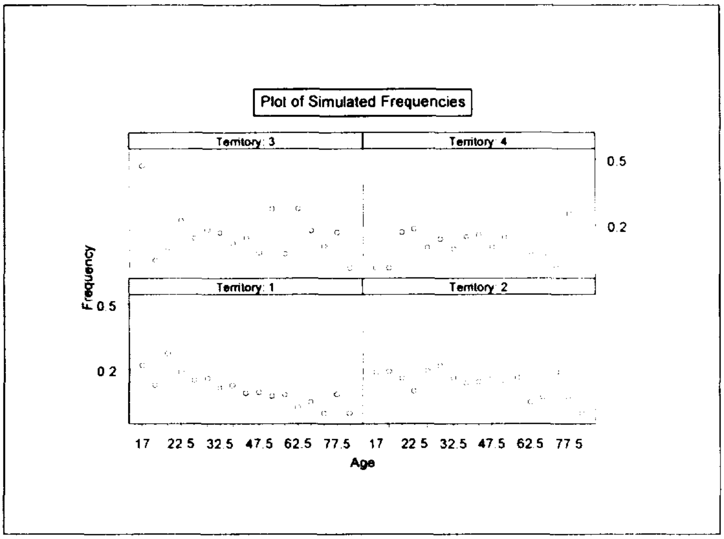
A simple approach to performing logistic regression (Hosmer and Lemshow, 1989), and the one which will be used for this paper, is to apply a weighted regression technique to aggregated data. This is done as follows:

1. Group the policyholder's into age groups such as 16 to 20, 21 to 25, etc.
2. Aggregate the claim counts and exposure counts (here the exposure is policyholders) by age group and territory.
3. Compute the frequency for each age and territory combination by dividing the number of claims by the number of policyholders.
4. Apply the logit transform to the frequencies (for logistic regression). That is compute $\log(p/(1-p))$ where p is the claim frequency or propensity. It may be necessary to add a very small quantity to the frequencies before the transform is computed, because some of the cells may have a frequency of 0.
5. Compute a value for driver age in each cell. The age data has been grouped and a value representative of driver ages in the cell is needed as an independent variable in the modeling. Candidates are the mean and median ages in the cell. The simplest approach is to use the midpoint of the age interval.
6. The policyholder count in each cell will be used as the weight in the regression. This has the effect of causing the regression to behave as if the number of observations for each cell equals the number of policyholders.

One of the advantages of using the aggregated data is that some observations have more than one claim. That is, the observations on individual records are not strictly binary, since values of 2 claims and even 3 claims sometimes occur. More complicated methods such as multinomial logistic regression[11] can be used to model discrete variables with more than 2 categories. When the data is aggregated, all the observations of the dependent variable are still in the range 0 to 1 and the logit transform still is appropriate for such data. Applying the logit transform to the aggregated data avoids the need for a more complicated approach. No transform was applied to the data to which the neural network was applied, i.e., the dependent variable was the observed frequencies. The result of aggregating the simulated data is displayed in Figure 29.

[11] A Poisson regression using Generalized Linear Models could also be used.

Figure 29



<u>Neural Network Results</u>

A five node neural network was fit to the data. The weights between the input and hidden layers are displayed in Table 11. If we examine the weights between the input and the hidden nodes, no variables seem insignificant, but it is hard to determine the impact that each variable is having on the result. Note that weights are not produced for Territory 4. This is the base territory in the neural network procedure and its parameters are incorporated into w_0, the constant.

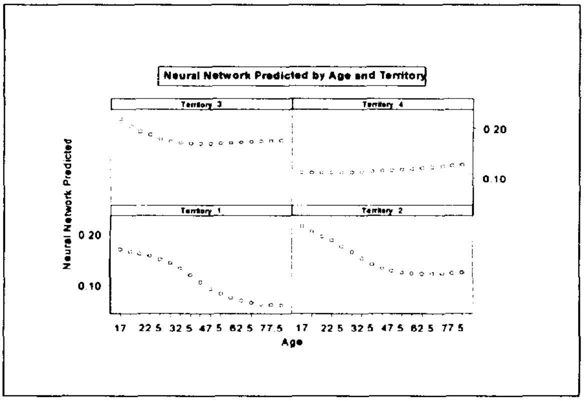
Table 11: Weights to Hidden Layer					
Node	W_0(Constant)	Weight(Age)	Weight(Territory 1)	Weight(Territory 2)	Weight(Territory 3)
1	-0.01	0.18	-0.02	-0.06	0.09
2	0.35	-0.01	-1.06	-0.73	-0.10
3	-0.36	0.21	-0.07	-0.82	0.46
4	-0.01	0.19	-0.01	-0.08	0.09
5	0.56	-0.08	-0.90	-1.10	-0.98

Interpreting the neural network is more complicated than interpreting a typical regression. In the previous section, it was shown that each variable's importance could be measured by a sensitivity. Looking at the sensitivities in Table 12, it is clear that both age and territory have a significant impact on the result. The magnitude of their effects seems to be roughly equal

Table 12: Sensitivity of Variables in Interaction Example	
Variable	Sensitivity
Age	24%
Territory	23%

Neither the weights nor the sensitivities help reveal the form of the fitted function. However graphical techniques can be used to visualize the function fitted by the neural network. Since interactions are of interest, a panel graph showing the relationship between age and frequency for each territory can be revealing. A panel graph has panels displaying the plot of the dependent variable versus an independent variable for each value of a third variable, or for a selected range of values of a third variable. (Examples of panel graphs have already been used in this paper in this section, to help visualize interactions). This approach to visualizing the functional form of the fitted curve can be useful when only a small number of variables are involved. Figure 30 displays the neural network predicted values by age for each territory. The fitted curve for territories 2 and 3 are a little different, even though the "true" curves are the same. The curve for territory 4 is relatively flat, although it has a slight upward slope.

Figure 30



Regression fit

Table 13 presents the fitted coefficients for the logistic regression. Interpreting these coefficients is more difficult than interpreting those of a linear regression, since the logit represents the log of the odds ratio *(p/(1-p))*, where *p* represents the underlying true claim frequency. Note that as the coefficients of the logit of frequency become more positive, the frequencies themselves become more positive. Hence, variables with positive coefficients are positively related to the dependent variable and coefficients with negative signs are negatively related to the dependent variable.

Table 13: Results of Regression Fit

Variable	Coefficient	Significance
Intercept	-1.749	0
Age	-0.038	0.339
Territory 1	-0.322	0.362
Territory 2	-0.201	0.451
Territory 3	-0.536	0.051
Age*Territory 1	0.067	0.112
Age*Territory 2	0.031	0.321
Age*Territory 3	0.051	0.079

Figure 31 displays the frequencies fitted by the logistic regression. As with neural networks graph are useful for visualizing the function fitted by a logistic regression. A noticeable departure from the underlying values can be seen in the results for Territory 4. The fitted curve is upward sloping for Territory 4, rather than flat as the true values are.

Figure 31

	Table 14	
Results of Fits:	Mean squared error	
	Training Data	Test Data
Neural Network	0.005	0.014
Regression	0.007	0.016

In this example the neural network had a better performance than the regression. Table 14 displays the mean square errors for the training and test data for the neural network and the logistic regression. Overall, the neural network had a better fit to the data and did a better job of capturing the interaction between Age and Territory. The fitted neural network model explained 30 % of the variance in the training data versus 15% for the regression. It should be noted that neither technique fit the "true" curve as closely as the curves in previous examples were fit. This is a result of the noise in the data. As can be seen from Figure 29, the data is very noisy, i.e., there is a lot of randomness in the data relative to the pattern. The noise in the data obscures the pattern, and statistical techniques applied to the data, whether neural networks or regression will have errors in their estimated parameters.

Example 5: An Example with Messy Data

The examples used thus far were kept simple, in order to illustrate key concepts about how neural networks work. This example is intended to be closer to the typical situation where data is messy. The data in this example will have nonlinearities, interactions, correlated variables as well as missing observations.

To keep the example realistic, many of the parameters of the simulated data were based on information in publicly available databases and the published literature. A random sample of 5,000 claims was simulated. The sample represents 6 years of claims history. (A multiyear period was chosen, so that inflation could be incorporated into the example). Each claim represents a personal automobile claim severity developed to ultimate[12]. As an alternative to using claims developed to ultimate, an analyst might use a database of claims which are all at the same development age. Random claim values were generated from a lognormal distribution. The scale parameter, μ, of the lognormal, (which is the mean of the logged variables) varied with the characteristics of the claim. The claim characteristics in the simulation were generated by eight variables. The variables are summarized in Table 15. The μ parameter itself has a probability distribution. A graph of the distribution of the parameter in the simulated sample is shown in Figure 32. The parameter had a standard deviation of approximately .38. The objective of the analysis is to distinguish high severity policyholders from low severity

[12] The analyst may want to use neural network or other data mining techniques to develop the data.

policyholders. This translates into an estimate of μ which is as close to the "true" μ as possible.

Figure 32

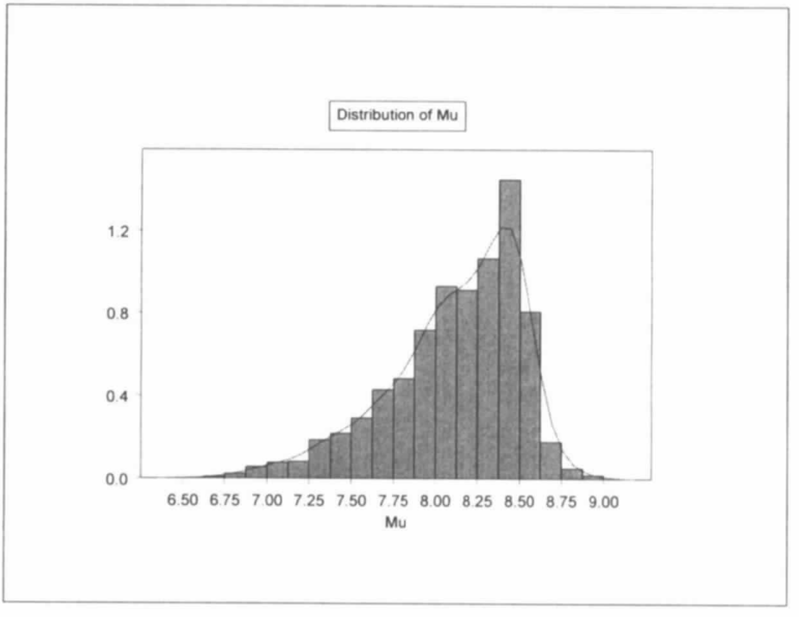


Table 15 below lists the eight predictor variable used to generate the data in this example. These variables are not intended to serve as an exhaustive list of predictor variables for the personal automobile line. Rather they are examples of the kinds of variables one could incorporate into a data mining exercise. A ninth variable (labeled Bogus) has no causal relationship to average severity. It is included as a noise variable to test the statistical procedures in their effectiveness at using the data. An effective prediction model should be able to distinguish between meaningful variables and variables which have no relationship to the dependent variable. Note that in the analysis of the data, two of the variables used to create the data are unavailable to the analyst as they represent unobserved variables (the Auto BI and Auto PD underlying inflation factors). Instead, six inflation indices which are correlated with the unobserved Factors are available to the analyst for modeling. Some features of the variables are listed below.

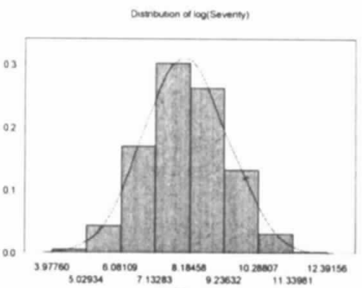
Table 15			
Variable	**Variable Type**	**Number of Categories**	**Missing Data**
Age of Driver	Continuous		No
Territory	Categorical	45	No
Age of Car	Continuous		Yes
Car Type	Categorical	4	No
Credit Rating	Continuous		Yes
Auto BI Inflation Factor	Continuous		No
Auto PD and Phys Dam Inflation Factor	Continuous		No
Law Change	Categorical	2	No
Bogus	Continuous		No

Note that some of the data is missing for two of the variables. Also note that a law change was enacted in the middle of the experience period which lowered expected claim severity values by 20%. A more detailed description of the variables is provided in Appendix 2.

Neural Network Analysis of Simulated Data

The dependent variable for the model fitting was the log of severity. A general rule in statistics is that variables which show significant skewness should be transformed to approximate normality before fitting is done. The log transform is a common transform for accomplishing this. In general, Property and Casualty severities are positively skewed. The data in this example have a skewness of 6.43, a relatively high skewness. Figure 33, a graph of the distribution of the log of severity indicates that approximate normality is attained after the data is logged.

Figure 33



The data was separated into a training database of 4,000 claims and a test database of 1,000 claims. A neural network with 7 nodes in the hidden layer was run on the 4,000 claims in the training database. As will be discussed later, this network was larger than the final fitted network. This network was used to rank variables in importance and eliminate some variables. Because the amount of variance explained by the model is relatively small (8%), the sensitivities were also small. Table 16 displays the results of the sensitivity test for each of the variables. These rankings were used initially to eliminate two variables from the model: Bogus, and the dummy variable for car age missing. Subsequent testing of the model resulted in dropping other variables. Despite their low sensitivities, the inflation variables were not removed. The low sensitivities were probably a result of the high correlations of the variables with each other. In addition, it was deemed necessary to include a measure of inflation in the model. Since the neural network's hidden layer performs dimension reduction on the inflation variables, in a manner analogous to Factor or Principal Components Analysis, it seemed appropriate to retain these variables.

Table 16: Sensitivities of Neural Network

Variable	Sensitivity	Rank
Car age	9.0	1
Age	5.3	2
Car type	3.0	3
Law Change	2.2	4
Credit category	2.2	5
Territory	2.0	6
Credit score	1.0	7
Medical Inflation	0.5	8
Car age missing	0.4	9
Hospital Inflation	0.1	10
Wage Inflation	0.0	11
Other Services Inflation	0.0	12
Bogus	0.0	13
Parts Inflation	0.0	14
Body Inflation	0.0	15

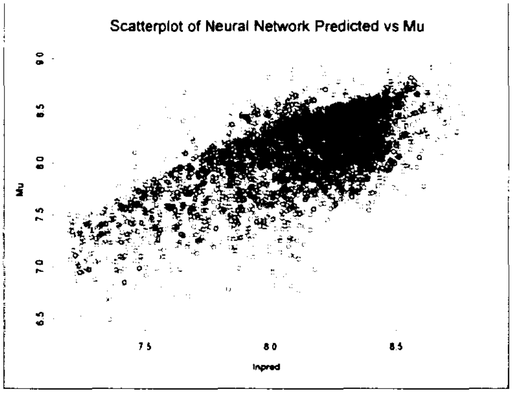
One danger that is always present with neural network models is overfitting. As more hidden layers nodes are added to the model, the fit to the data improves and the r^2 of the model increases. However, the model may simply be fitting the features of the training data, therefore its results may not generalize well to a new database. A rule of thumb for the number of intermediate nodes to include in a neural network is to use one half of the number of variables in the model. After eliminating 2 of the variables, 13 variables remained in the model. The rule of thumb would indicate that 6 or 7 nodes should be used. The test data was used to determine how well networks of various sizes performed when presented with new data. Neural networks were fit with 3, 4, 5, 6 and 7 hidden nodes. The fitted model was then used to predict values of claims in the test data. Application of the fitted model to the test data indicated that a 4 node neural network provided the best model. (It produced the highest r^2 in the test data). The test data was also used to eliminate additional variables from the model. In applying the model to the test data it was found that dropping the territory and credit variables improved the fit.

Goodness of Fit

The fitted model had an r^2 of 5%. This is a low r^2 but not out of line with what one would expect with the highly random data in this example. The "true" μ (true expected log (severity)) has a variance equal to 10% of the variance of the log of severity. Thus, if one had perfect knowledge of μ, one could predict individual log(severities) with only 10% accuracy. However, if one had perfect knowledge of the true mean value for severity for each policyholder, along with knowledge of the true mean frequency for each policyholder, one could charge the appropriate rate for the policy, given the particular characteristics of the policyholder. In the aggregate, with a large number of policyholders, the insurance company's actual experience should come close to the experience predicted from the expected severities and frequencies.

With simulated data, the "true" μ for each record is known. Thus, the model's accuracy in predicting the true parameter can be assessed. Figure 34 plots the relationship between μ and the predicted values (for the log of severity). It can be seen that as the predicted value increases, μ increases. The correlation between the predicted values and the parameter mu is .7.

Figure 34



As a further test of the model fit, the test data was divided into quartiles and the average severity was computed for each quartile. A graph of the result is presented in Figure 35. This graph shows that the model is effective in discriminating high and low severity claims. One would expect an even better ability to discriminate high severity from low severity observations with a larger sample. This is supported by Figure 36 which displays the plot of "true" expected severities for each of the quartiles versus the neural network predicted values. This graph indicated that the neural network is effective in classifying claims into severity categories. These results suggest that neural networks could be used to identify the more profitable insureds (or less profitable insureds) as part of the underwriting process.

Figure 35

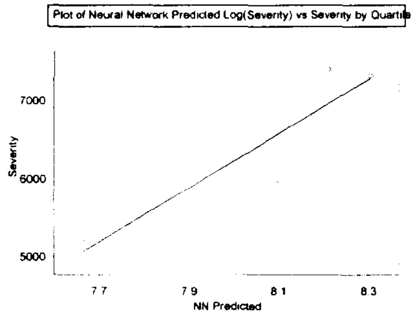
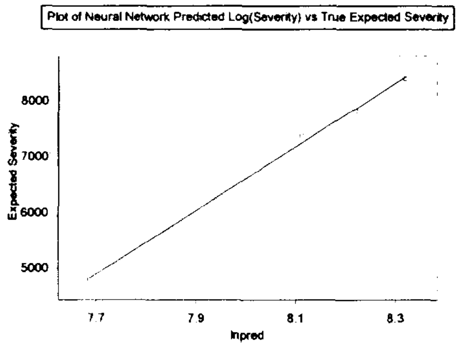


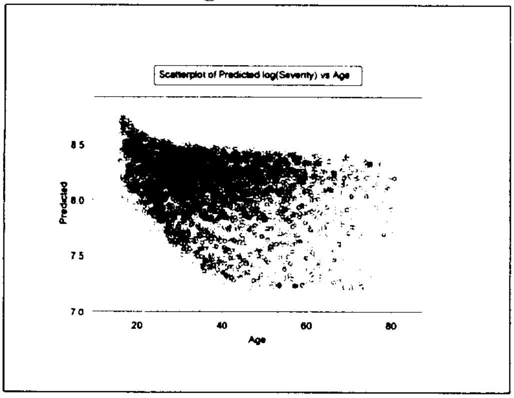
Figure 36



<u>Interpreting Neural Networks Revisited: Visualizing Neural Network Results</u>

In the previous example some simple graphs were used to visualize the form of the fitted neural network function. Visualizing the nature of the relationships between dependent and independent variables is more difficult when a number of variables are incorporated into the model. For instance, Figure 37 displays the relationship between the neural network predicted value and the driver's age. It is difficult to discern the relationship between age and the network predicted value from this graph. One reason is that the predicted value at a given age is the result of many other predictor variables as well as age. Thus, there is a great deal of dispersion of predicted values at any given age due to these other variables, disguising the fitted relationship between age and the dependent variable.

Figure 37



Researchers on neural networks have been exploring methods for understanding the function fit by a neural network. Recently, a procedure for visualizing neural network fitted functions was published by Plate, Bert and Band (Plate et al., 2000). The procedure is one approach to understanding the relationships being modeled by a neural network. Plate et al. describe their plots as Generalized Additive Model style plots. Rather than attempting to describe Generalized Additive Models, a technique for producing the plots is simply presented below. (Both Venables and Ripley and Plate et al. provide descriptions of Generalized Additive Models). The procedure is implemented as follows:

1. Set all the variables except the one being visualized to a constant value. Means and medians are logical choices for the constants.
2. Apply the neural network function to this dataset to produce a predicted value for each value of the independent variable. Alternatively, one could choose to apply the neural network to a range of values for the independent variable selected to represent a reasonable set of values of the variable. The other variables remain at the selected constant values.

3. Plot the relationship between the neural network predicted value and the variable.
4. Plate et al. recommend scaling all the variables onto a common scale, such as 0 to 1. This is the scale of the inputs and outputs of the logistic functions in the neural network. In this paper, variables remain in their original scale.

The result of applying the above procedure is a plot of the relationship between the dependent variable and one of the independent variable. Multiple applications of this procedure to different variables in the model provides the analyst with a tool for understanding the functional form of the relationships between the independent and dependent variables.

The visualization method was applied to the data with all variables set to constants except for driver age. The result is shown in Figure 38. From this graph we can conclude that the fitted function declines with driver age. Figure 39 shows a similar plot for car age. This function declines with car age, but then increases at older ages.

Figure 38

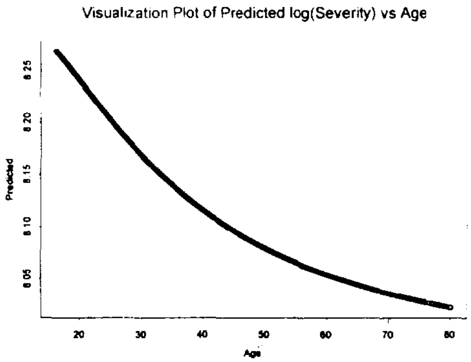
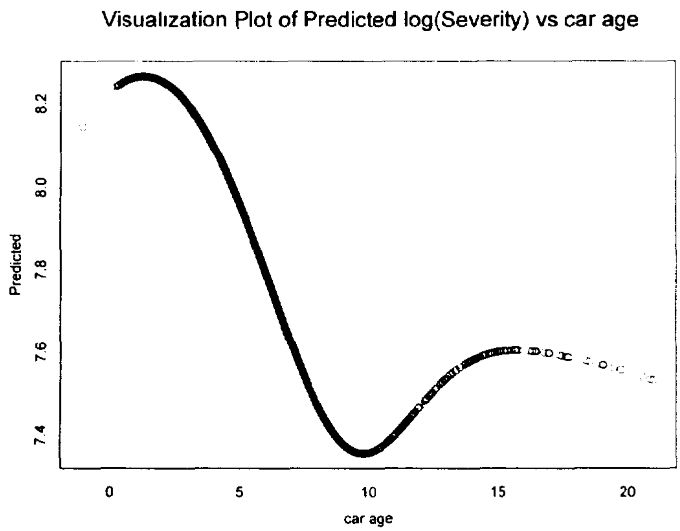
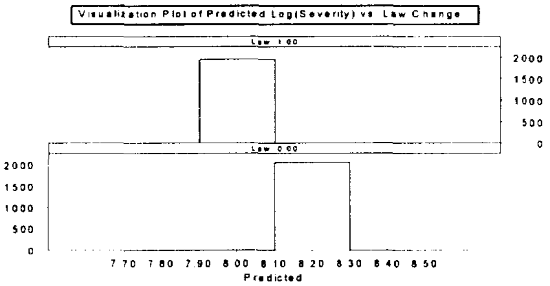


Figure 39



Suppose we wanted to visualize the relationship between a predictor variable which takes on discrete values and the dependent variable. For instance, suppose we wanted to know the impact of the law change. We can create fitted values for visualizing as described above but instead of producing a scatterplot, we can produce a bar chart. Figure 40 displays such a graph. On this graph, the midpoint for claims subject to the law change (a value of 1 on the graph) is about .2 units below the midpoint of claims not subject to the law change. This suggests that the neural network estimates the law effect at about 20% because a .2 impact on a log scale corresponds approximately to a multiplicative factor of 1.2, or .8 in the case of a negative effect (Actually, the effect when converted from the log scale is about 22%). The estimate is therefore close to the "true" impact of the law change, which is a 20% reduction in claim severity.

Figure 40



The visualization procedure can also be used to evaluate the impact of inflation on the predicted value. All variables except the six economic inflation factors were fixed at a constant value while the inflation variables entered the model at their actual values. The predicted values are then plotted against time. Figure 41 shows that the neural network estimated that inflation increased by about 40% during the six year time period of the sample data. This corresponds roughly to an annual inflation rate of about 7%. The "true" inflation underlying the model was approximately 6%.

One way to visualize two-way interactions is to allow two variables to take on their actual values in the fitting function while keeping the others constant. Figure 42 displays such a panel graph for the age and car age interaction. It appears from this graph that the function relating car age to the predicted variable varies with the value of driver age.

Figure 41

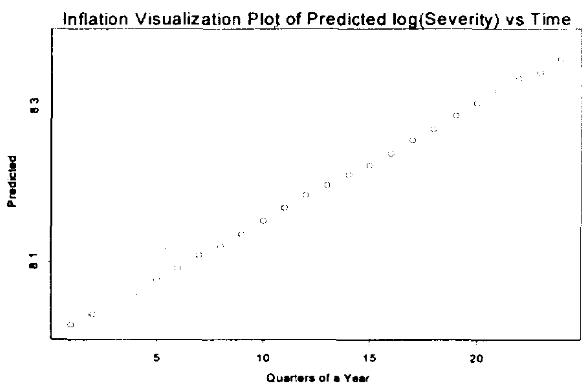
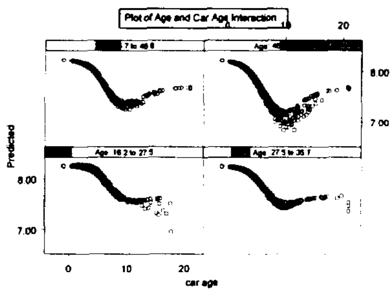


Figure 42



Regression Model

A regression model was fit to the data. The dependent variable was the log of severity. Like neural networks, regression models can be subject to overfitting. The more variables in the model, the better the fit to the training data. However, if the model is overfit it will not generalize well and will give a poor fit on new data. Stepwise regression is an established procedure for selecting variables for a regression model. It tests the variables to find the one that produces the best r^2. This is added to the model. It continues cycling through the variables, testing variables and adding a variable each cycle to the model until no more significant variables can be found. Significance is usually determined by performing an F-test on the difference in the r^2 of the model without a given variable and then with the variable.

Stepwise regression was used to select variables to incorporate into the model. Then a regression on those variables was run. The variables selected were driver age, car age, a dummy variable for the law change and the hospital inflation factor. Note that the hospital inflation factor had a very high correlation with both underlying inflation factors (even though the factors were generated to be independent of each other[13]). Thus, using just the one variable seems to adequately approximate inflation. On average, the increase in the hospital inflation index was 4.6%. Since a factor of 1.15 (see Table 17) was applied to the hospital inflation factor, inflation was estimated by the regression to be a little over 5% per year. The regression model estimated the impact of the law change as a reduction of .3 on the log scale or about 35% as opposed to the estimate of about 22% for the neural network. Thus, the neural network overestimated inflation a little, while the regression model underestimated it a little. The neural network estimate of the law

[13] This may be a result of using a random walk procedure to generate both variable. Using random walk models, the variables simulated have high correlations with prior values in the series.

change effect was close to the "true" value, while the regression overestimated the magnitude of the effect.

The regression found a negative relationship between driver age and severity and between car age and severity. An interaction coefficient between age and car age was also estimated to be negative. The results correspond with the overall direction of the "true" relationships. The results of the final regression are presented in Table 17.

The fitted regression had a somewhat lower r^2 than the neural network model. However, on some goodness of fit measures, the regression performance was close to that of the neural network. The regression predicted values had a .65 correlation with μ. versus .70 for the neural network. As seen in Figures 43 and 44, the regression was also able to discriminate high severity from low severity claims with the test data. Note that neither model found the Bogus variable to be significant. Also, neither model used all the variables that were actually used to generate the data, such as territory or credit information. Neither technique could distinguish the effect of these variables from the overall background noise in the data.

Table 17: Regression Results

Variable	Coefficient	Significance
Intercept	7.210	0
Age	-0.001	0.448
car age	-0.024	0.203
Law	-0.306	0.0001
Hospital Inf	1.1	0.0059
Age*car age	-0.001	0.0195
R^2 = .039		

Figure 43

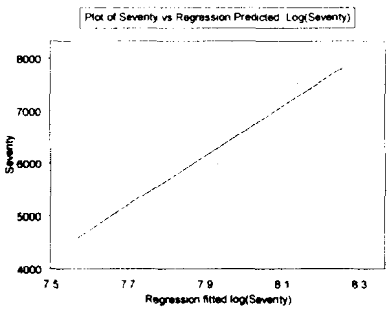
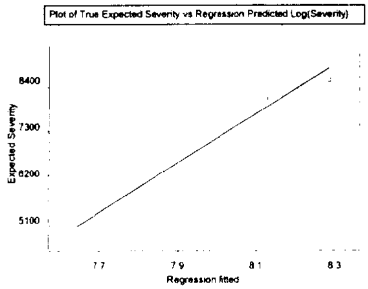


Figure 44



Using the model in prediction

To estimate severities, the fitted log severities must be transformed back to their original scale. This is generally accomplished by applying the exponential function to the values predicted by the model. If the data is approximately lognormally distributed, as in this example, a simple exponential transform will understate the true value of the predicted severity. The mean of a lognormal variable is given by:

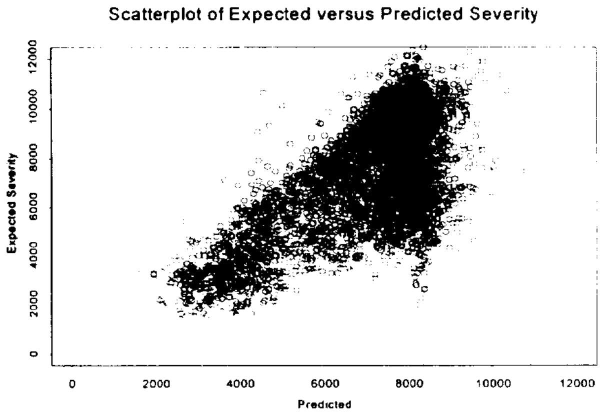
$$E(Y_i) = e^{\mu_i + .5*\sigma^2}$$

where

$E(Y_i)$ is the expected value for the i^{th} observation
μ_i = the mean for i^{th} observation on the log scale
σ^2 is the variance of severities on a log scale

Since μ_i and σ^2 are unknown, estimates of their values must be used. The predicted value from the neural network or regression is the usual choice for an estimate of μ_i. The mean square error of the neural network or regression can be used as an estimate of σ^2 in the formula above. A predicted value was computed for the claims that were used to fit the neural network model. A plot of the predicted severities versus the "true" expected severities is displayed in Figure 43.

Figure 43



<u>Applying the models</u>
Some of the possible applications of neural networks and other modeling techniques can utilize predictions of claim severity. A company may want to devise an early warning system to screen newly reported claims for those with a high probability of developing into large settlements. A severity model utilizing only information available early in the life of a claim could be used in an early warning system. A fraud detection system could also be based on claim severity. One approach to fraud detection is to produce a severity prediction for each claim. The actual value of the claim is compared to the predicted value. Those with a large positive deviation from the predicted are candidates for further investigation.

However, many of the underwriting applications of modeling and prediction require both a frequency and a severity estimate. A company may wish to prune "bad" risks from its portfolio, pursue "good" risks or actually use models to establish rates. For such applications either the loss ratio or pure premium will be the target variable of interest. There are two approaches to estimating the needed variable: 1) One can separately estimate frequency and severity models and combine the estimates of the two models. An illustration of fitting models to frequencies was provided in Example 4 and an example of fitting models to severities was supplied in Example 5. 2) Alternatively, one can estimate a pure premium or loss ratio model directly.

One difficulty of modeling pure premiums or loss ratios is that in some lines of business, such as personal lines, most of the policyholders will have no losses, since the expected frequency is relatively low. It is desirable to transform the data onto a scale that does not allow for negative values. The log transformation accomplishes this. However, since the log is not defined for a value of zero it may be necessary to add a very small constant to the data in order to apply the log transform.

Once a pure premium is computed, it can be converted into a rate by loading for expenses and profit. Alternatively, the pure premium could be ratioed to premium at current rate levels to produce a loss ratio. A decision could be made as to whether the predicted loss ratio is acceptable before underwriting a risk. Alternatively the loss ratio prediction for a company's portfolio of risks for a line of business can be loaded for expenses and profit and the insurance company can determine if a rate increase is needed.

Summary

This paper has gone into some detail in describing neural networks and how they work. The paper has attempted to remove some of the mystery from the neural network "black box". The author has described neural networks as a statistical tool which minimizes the squared deviation between target and fitted values, much like more traditional statistical procedures do. Examples were provided which showed how neural networks 1) are universal function approximators and 2) perform dimension reduction on correlated predictor variables. Classical techniques can be expected to outperform neural network models when data is well behaved and the relationships are linear or variables can be transformed into variables with linear relationships. However neural networks seem to have an advantage over linear models when they are applied to complex nonlinear data. This is an advantage neural networks share with other data mining tools not discussed in detail in this paper. Future research might investigate how neural networks compare to some of these data mining tools.

Note that the paper does not advocate abandoning classical statistical tools, but rather adding a new tool to the actuarial toolkit. Classical regression performed well in many of the examples in this paper. Some classical statistical tools such as Generalized Linear Models have been applied successfully to problems similar to those in this paper. (See Holler et al. for an example).

A disadvantage of neural networks is that they are a "black box". They may outperform classical models in certain situations, but interpreting the result is difficult because the nature of the relationship between dependent and target variables is not usually revealed. Several methods for interpreting the results of neural networks were presented. Methods for visualizing the form of the fitted function were also presented in this paper. Incorporating such procedures into neural network software should help address this limitation.

Appendix 1: Neural Network Software

Neural network software is sold at prices ranging from a couple of hundred dollars to $100,000 or more. The more expensive prices are generally associated with more comprehensive data mining products, which include neural networks as one of the capabilities offered. Some of the established vendors of statistical software such as SPSS and SAS sell the higher end data mining products[14]. These products are designed to function on servers and networks and have the capability of processing huge databases. They also have some of the bells and whistles useful to the analyst in evaluating the function fit by the neural network, such as a computation of sensitivities. Both of these products allow the user to apply a number of different kinds of neural networks, including types of networks not covered in this paper.

Many of the less expensive products provide good fits to data when the database is not large. Since the examples in this paper used modestly sized databases, an expensive product with a lot of horsepower was not required. Two of the less expensive tools were used to fit the models in this paper: a very inexpensive neural network package, Brainmaker, and the S-PLUS neural network function, nnet. The Brainmaker tool has a couple of handy features. It creates a file that contains all the parameters of the fitted neural network function for the hidden and output layers. It also has the capability of producing the values of the hidden nodes. Both of these features were helpful for the detailed examination of neural networks contained in this paper. However, the Brainmaker version employed in this analysis had difficulty fitting networks on larger databases[15], so the S-PLUS nnet function was used for the last example. The S-PLUS nnet function is contained in a library supplied by Venables and Ripley, rather than the vendors of S-PLUS, but it is included in the basic S-PLUS package. This software also provides the fitted parameters for the hidden and output layers. (However, it does not provide the fitted values for the hidden nodes). Chapter 9 of Venables and Ripley describes the software and how to use it. (Venables and Ripley, 1999).

The commonly used commercial software for fitting neural networks does not incorporate the visualization technique used for Example 5. Plate has provided an S-PLUS library incorporating his visualization technique (which is similar to, but a little different from, the one used for this paper) in the statlib library at http://lib.stat.cmd.edu/S/. The library with the visualization software is named Ploteff.

Numerous other products with which the author of this paper has no experience are also available for fitting neural networks. Thus, no statement made in this paper should be interpreted as an endorsement of any particular product.

---

[14] The SPSS data mining product is called Clementine. The SAS product is called the Enterprise Miner. SPSS also sells an inexpensive neural network product, Neural Connection. The author has used Neural Connection on moderately sized databases and found it to be effective on prediction and classification problems.

[15] It should be noted that the vendors of Brainmaker sell a professional version which probably performs better on large databases.

Appendix 2

This appendix is provided for readers wishing a little more detail on the structure of the data in the Example 5.

The predictor variables are:
Driver age: Age of the driver in years
Car type: This is intended to represent classifications like compact, midsize, sports utility vehicle and luxury car. There are 4 categories.
Car age: Age of the car in years
Representative parameters for the Driver age, Car type and Car age and their interactions variables were determined from the Baxter automobile claims database[16]
Territory: Intended to represent all the territories for 1 state. There are 45 categories. Reasonable parameters for territory were determined after examining the Texas automobile database used in the Casualty Actuarial Society's ratemaking competition.
Credit: A variable called leverage, representing the ratio of the sum of all revolving debt to the sum of all revolving credit limits was used as an indicator of the creditworthiness of the driver. This is a variable not typically used in ratemaking. However, some recent research has suggested it may be useful in predicting personal lines loss ratios. Monaghan (Monagahan, 2000) shows that credit history has a significant impact on personal automobile and homeowners' loss ratios. Monaghan discussed a number of possible credit indicators, which were useful in predicting loss ratios. The leverage variable was judgmentally selected for this model because it had high predictive accuracy and because parameters could be developed based on information in Monaghan's paper. If a company had access to its policyholders' credit history, it might wish to develop a separate credit score (perhaps using neural networks) which used the information of a number of credit history variables. Another credit variable was used in addition to the leverage ratio. People with a leverage ratio of 0 were divided into 2 categories, those with very low limits (< $500) and those with higher limits (>=$500). A third category was created for claimants with leverage greater than 0. For the purposes of illustrating this technique, it was assumed that the entire impact of the credit variable is on severity, although this is unlikely in practice.

Automobile Bodily Injury (ABI) inflation factor and Automobile Property Damage and Physical Damage (APD) inflation factor: These factors drive quarterly increases in the bodily injury, property damage and physical damage components of average severity. They are unobserved factors. The ABI factor is correlated with three observed variables: the producer price index for hospitals, the medical services component of the consumer price index and an index of average hourly earnings. The APD factor is correlated with three observed variables: the produce price index for automobile bodies, the producer price index for automobile parts and the other services component of the consumer price index. Bureau of Labor statistics data was reviewed when developing parameters for the factors and for the "observed" variables. The ABI factor was given a 60% weight and the APD factor was given a 40% weight in computing each claim's expected severity.

---

[16] This database of Automobile claims is available as an example database in S-PLUS. Venables and Ripley supply the S-PLUS data for claim severities in a S-PLUS library. See Venables and Ripley, p.467.

Law Change: A change in the law is enacted which causes average severities to decline by 20% after the third year.

Interactions:
Table 18 shows the variables with interactions. Three of the variables have interactions. In addition some of the interactions are nonlinear (or piecewise linear). An example is the interactions between age and car age. This is a curve that has a negative slope at older car ages and younger driver ages, but is flat for older driver ages and younger car ages. The formula used for generating the interaction between age, car age and car type is provided below (after Table 19). In addition to these interactions, other relationships exist in the data, which affect the mix of values for the predictor variables in the data. Young drivers (<25 years old) are more likely not to have any credit limits (a condition associated with a higher average severity on the credit variable). Younger and older (>55) drivers are more likely to have older cars.

Table 18

Interactions
Driver Age and Car Type
Driver Age and Car Age
Driver Age and Car Age and Car Type

Nonlinearities
A number of nonlinear relationships were built into the data. The relationship between Age and severity follows an exponential decay (see formula below). The relationships between some of the inflation indices and the Factors generating actual claim inflation are nonlinear. The relationship between car age and severity is piecewise linear. That is, there is no effect below a threshold age, then effect increases linearly up to a maximum effect and remains at that level at higher ages.

Missing Data
In our real life experience with insurance data, values are often missing on variables which have a significant impact on the dependent variable. To make the simulated data in this example more realistic, data is missing on two of the independent variables. Table 19 presents information on the missing data. Two dummy variables were created with a value of 0 for most of the observations, but a value of 1 for records with a missing value on car age and/or credit information. In addition, a value of −1 was recorded for car age and credit leverage where data was missing. These values were used in the neural network analysis. The average of each of the variables was substituted for the missing data in the regression analysis.

Table 19	
Missing Values	
Car Age	10% of records missing information if driver age is < 25. Otherwise 5% of data is missing
Credit	25% of records are missing the information if Age < 25, otherwise 20% of data is missing.

The μ parameter of the lognormal severity distribution was created with the following function:

BF1 = max(0, min(4,carage[I] - 6))
BF2= (cartype[I] = 4 or cartype[I]=2)
BF3= max(0, min(6,(carage[I] – 3)))
BF4= (cartype[I] = 1 or cartype[I] = 3 or cartype[I] = 4) * BF3

μ [I]<-(7.953)-.05*BF1+ 2* exp(-.15*Age[I])*BF4* exp(-.15*Age[I])*BF3 -0.15 * BF3+ 1.5*exp(-.1*Age[I])* BF4+log(terrfactor)+Law[I]*log(.8)+log(leverage[I]))+log(Factor1*.6) + log(Factor2*.4)

where

I is the index of the simulated observation
BF1, BF2, BF3, BF4 are basis functions which are used to incorporate interaction variables and piecewise linear functions into the function for μ.

μ[I] is the lognormal mu parameter for the i^{th} record
Age is the driver's age
cartype is the car type.
carage is the car's age
terrfactor is the multiplicative factor for territory
Law is an indicator variable, which is 0 for quarters 1 through 12 and 1 afterwards.
leverage is the multiplicative factor for the claimant's credit leverage
Factor1, Factor2 are the bodily injury and property damage inflation factors.

The dispersion parameter of the lognormal, σ, was 1.2.

References

Berry, Michael J. A., and Linoff, Gordon, *Data Mining Techniques*, John Wiley and Sons, 1997

Dhar, Vasant and Stein, Roger, *Seven methods for Transforming Corporate Data Into Business Intelligence*, Princeton Hall, 1997

Dunteman, George H., *Principal Components Analysis*, SAGE Publications, 1989

Derrig, Richard, "Patterns, Fighting Fraud With Data", *Contingencies*, pp. 40–49.

Freedman, Roy S., Klein, Robert A. and Lederman, Jess, *Artificial Intelligence in the Capital Markets*, Probus Publishers 1995

Hatcher, Larry, *A Step by Step Approach to Using the SAS System for Factor Ananlysis*, SAS Institute, 1996

Heckman, Phillip E. and Meyers, Glen G., "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Cost Distributions", *Proceedings of the Casualty Actuarial Society*, 1983, pp. 22-61.

Holler, Keith, Somner, David, and Trahair, Geoff, "Something Old, Something New in Classification Ratemaking With a New Use of GLMs for Credit Insurance", *Casualty Actuarial Society Forum, Winter 1999, pp. 31-84.*

Hosmer, David W. and Lemshow, Stanley, *Applied Logistic Regression*, John Wiley and Sons, 1989

Keefer, James, "Finding Causal Relationships By Combining Knowledge and Data in Data Mining Applications", Paper presented at Seminar on Data Mining, University of Delaware, April, 2000.

Kim, Jae-On and Mueler, Charles W, *Factor Analysis: Statistical Methods and Practical Issues*, SAGE Publications, 1978

Lawrence, Jeannette, *Introduction to Neural Networks: Design, Theory and Applications*, California Scientific Software, 1994

Martin, E. B. and Morris A. J., "Artificial Neural Networks and Multivariate Statistics", in *Statistics and Neural Networks: Advances at the Interface*, Oxford University Press, 1999, pp. 195 – 292

Masterson, N. E., "Economic Factors in Liability and Property Insurance Claims Cost: 1935 – 1967", *Proceedings of the Casualty Actuarial Society*, 1968, pp. 61 – 89.

Monaghan, James E., "The Impact of Personal Credit History on Loss Performance in Personal Lines", *Casualty Actuarial Society Forum*, Winter 2000, pp. 79-105

Plate, Tony A., Bert, Joel, and Band, Pierre, "Visualizing the Function Computed by a Feedforward Neural Network", *Neural Computation*, June 2000, pp. 1337-1353.

Potts, William J.E., *Neural Network Modeling: Course Notes*, SAS Institute, 2000

SAS Institute, *SAS/STAT Users Guide: Release 6.03*, 1988

Smith, Murry, *Neural Networks for Statistical Modeling*, International Thompson Computer Press, 1996

Speights, David B, Brodsky, Joel B., Chudova, Durya I.,, "Using Neural Nwtworks to Predict Claim Duration in the Presence of Right Censoring and Covariates", *Casualty Actuarial Society Forum*, Winter 1999, pp. 255-278.

Venebles, W.N. and Ripley, B.D., Modern *Applied Statistics with S-PLUS*, third edition, Springer, 1999

Warner, Brad and Misra, Manavendra, "Understanding Neural Networks as Statistical Tools", *American Statistician*, November 1996, pp. 284 - 293

Actuarial Applications of Multifractal Modeling Part I: Introduction and Spatial Applications

John A. Major, ASA, MAAA and
Yakov Lantsman, Ph.D.

Actuarial Applications of Multifractal Modeling

Part I: Introduction and Spatial Applications

by John A. Major, ASA, MAAA and Yakov Lantsman, Ph.D.[1]

email:jmajor@guycarp.com, lantsman@netrisk.com

Abstract

Multifractals are mathematical generalizations of fractals, objects displaying "fractional dimension," "scale invariance," and "self-similarity." Many natural phenomena, including some of considerable interest to the casualty actuary (meteorological conditions, population distribution, financial time series), have been found to be well-represented by (random) multifractals. In this part I paper, we define and characterize multifractals and show how to fit and simulate multifractal models in the context of two-dimensional fields. In addition, we summarize original research we have published elsewhere concerning the multifractal distribution of insured property values, and discuss how we have used those findings in particular and multifractal modeling in general in a severe storm catastrophe model.

Introduction

In this section, we introduce the concepts of fractals and multifractals.

Fractals

Mathematicians have known of sets whose dimension is not a whole number for some time, but the term "fractal" emerged on the scientific and popular scenes with the work of Benoit Mandlebrot in the 1960s and 1970s [Mandlebrot 1982].

Mathematically, a fractal can be defined as a point set with possibly non-integer dimension. Examples of fractals include continuous random walks (Weiner processes), the Cantor set, and the Sierpinski triangle (the latter two discussed below). Phenomena in natųre that resemble fractals include dust spills and coastlines.

Regular fractals possess the attribute of *self-similarity*. This means that parts of the set are similar (in the geometrical sense of equivalence under a transformation consisting of magnification, rotation, translation, and reflection) to the whole. This gives regular fractals an "infinite regress" look, as the same large-scale geometrical features are repeated at ever smaller and smaller scales. Self-similarity is also known as *scale*

---

[1] The authors would like to thank John Mangano for his contributions to this paper, Shaun Lovejoy and Daniel Schertzer for their helpful conversations, and Gary Venter for his review of an early draft. Errors, of course, are solely the responsibility of the authors.

symmetry or *scaling* – the fractal doesn't have a characteristic scale at which its features occur; they occur at all scales equally.

Irregular fractals do not possess strict self-similarity, but possess *statistical* self-similarity and scaling. This will be clarified below.

The key numerical index of a fractal, its *fractal dimension*, deserves further explanation. It is not immediately obvious how the concept of dimension from linear algebra, the maximum number of linearly independent vectors in a space, can be generalized to include the possibility of noninteger values. While there are several ways of doing so – and they often coincide – the so-called capacity dimension (sometimes misnamed the Hausdorff dimension[2]) is perhaps the easiest to understand.

Consider a closed and bounded subset S of N-dimensional Euclidean space R^N. We define a covering of S of size λ to be a set of hypercubes $\{H_i\}$ such that (1) each hypercube is of size λ on a side and (2) the set S is contained within the union of all hypercubes $\cup H_i$. For any λ, let $n(\lambda)$ be the minimum number of hypercubes needed to be a covering of S. The dimension of S can then be defined in terms of the scaling behavior of coverings of S, i.e., the behavior of $n(\lambda)$ as $\lambda \to 0$.

Examples:

- If S consists of a finite number of points, then for all λ less than the minimum distance between points, a covering needs to have as many hypercubes as there are points: n is constant for small λ.
- If S consists of a line segment of length L, then $n(\lambda)=L/\lambda$: n varies as the reciprocal of the first power of scale λ.
- If S consists of a (sub-) hypercube of dimension m and length L on a side, then $n(\lambda)$ is approximately $(L/\lambda)^m$: n varies as the reciprocal of the mth power of scale λ.

This exponential relation, $n(\lambda) \propto \lambda^{-m}$, motivates the definition of fractal dimension:

$$d = -\lim_{\lambda \to 0}\left(\frac{\log(n(\lambda))}{\log(\lambda)}\right) \qquad (1)$$

The previous examples show that by this definition, a set of isolated points has dimension zero, a line segment has dimension one, and an m-hypercube has dimension m, as we would expect.[3]

Subsets of the unit interval may have various dimensions less than or equal to one, and cardinality is no guarantee of dimension for infinite sets. Finite point sets have

---

[2] The definition of Hausdorff dimension is more technically complicated, involving an infimum rather than a limit, thereby handling cases where the limit (in equation 1 below) does not exist.

[3] Note that the dimension N of the embedding space is irrelevant. While it is true that a line segment of finite length can be made to fit in a hypercube of arbitrarily small side if the dimension of the hypercube is big enough, what really matters is the scaling behavior. That is, if the side of the hypercube is halved, then two of them are needed to cover the line segment – implying the line segment has dimension one.

dimension zero, of course, but there are countable subsets with dimension zero and those with dimension one. For example, the set of rational numbers (a countable set) is a dense subset of the real numbers, meaning that any open set around a real number contains a rational. Therefore, the fractal dimension of the rationals is the same as that of the reals (they need exactly the same covering sets), that is, one.

On the other hand, the countable set consisting of points $x_k = \alpha^k$, k=1,2,3,... where $0<\alpha<1$, has dimension 0. This can be seen by considering coverings by blocks of length $\lambda = \alpha^j$ for some arbitrary j. The first block covers all points x_j, x_{j+1}, and at most (j-1) blocks are needed to cover the other (j-1) points. Thus,

$$\log(n(\lambda))/\log(\lambda) \leq \log(j)/(j*\log(\alpha)) \rightarrow 0.$$

Nothing in the definition of fractal dimension precludes the possibility of a set S having a noninteger dimension d. We now present some examples to show how this can happen.

The *Cantor set* is a subset of a line segment and is defined recursively as follows. Start with the entire line segment. Remove the middle third, leaving two disconnected closed line segments. Repeat the process on each remaining line segment, ad infinitum. In the limit, we have the Cantor set. At stage k of the construction (the whole segment being stage 0), we have 2^k subsegments each of length 3^{-k}, for a total length of $(2/3)^k$. In the limit, the Cantor set has measure[4] zero (it consists of points with no net length) because in the limit, $(2/3)^k$ goes to zero. For any length $\lambda=3^{-k}$, we need 2^k segments H_i to cover the set. Therefore the fractal dimension of the Cantor set is log(2)/log(3) = 0.63093..., corresponding to something between a line and a set of isolated points.

The self-similarity of the Cantor set follows directly from its construction. Each subsegment is treated in precisely the same way (up to a scale factor) as the original segment.

As an example of a noninteger fractal dimension in a 2-dimensional space, consider the *Sierpinski triangle* (also known as the Sierpinski *gasket*). This subset of the unit square is defined recursively as follows: Start with an equilateral triangle and its interior. Draw an inscribed triangle (point down) connecting the midpoints of each side. This divides the triangle into four similar and congruent sub-triangles. Remove the interior of the inscribed triangle. Repeat the process on each of the remaining three sub-triangles. Figure 1 shows an approximation to the result. As with the Cantor set, the Sierpinski triangle has zero measure (no area), because each stage of the construction takes up $(3/4)^k$ area of the outer triangle. Assuming the original triangle is inscribed in a unit square, at stage k of the construction, we need 3^k squares H_i of side $\lambda=2^{-k}$ to cover the set. Therefore, the Sierpinski triangle has fractal dimension log(3)/log(2) = 1.584963..., corresponding to something between a linear and a planar figure.

The self-similarity of the Sierpinski triangle again follows directly from its construction. Each sub-triangle is a miniature version of the original triangle and is similar to all other triangles appearing in the set.

---

[4] Measure theory is reviewed in the next section.

The analysis of fractal dimension by this method is generally termed *box-counting*. There are other approaches, but they will not be discussed here. Note that the method applies to arbitrary sets, not just self-similar ones. A non-self-similar set is called an *irregular* fractal if it has a noninteger fractal dimension.

Among natural phenomena, coastlines are frequently cited as good examples of irregular fractals. The measured length of a coastline depends on the scale of accuracy of the measuring tool. Comparing maps at various scales, one can see progressive deterioration of detail as larger scales are used. What appears as a wrinkled inlet on one map is abstracted to a simple polygon on the next and then obliterated completely on the next. [Barnsley] gives the fractal dimension of the coast of Great Britain as approximately 1.2. [Woo] discusses numerous areas where fractal laws relate to natural hazard processes.

This notion of scale-dependent measurements will play a central role in the practical application of fractal and multifractal theory to real-world problems.

Multifractals

Multifractals, also known as *fractal measures*, generalize the notion of fractals. Mandlebrot also worked on multifractals in the 1970s and 1980s [Mandlebrot 1988], but the first use of the term is credited to U. Frisch and G. Parisi [Mandlebrot 1989]. Rather than being point sets, multifractals are measures (distributions) exhibiting a spectrum of fractal dimensions.

A brief review of measure theory is in order. A measure μ on a space X is a function from a set of subsets of X (a σ-algebra of "measurable sets") to the real numbers R. In order to be a measure, the function μ must satisfy $\mu(\emptyset)=0$, $\mu(S)\geq 0$, and μ of any countable collection of disjoint sets must equal the sum of μ on each set. Actuaries typically encounter only *probability* measures, where, in addition, $\mu(X)=1$. The usual measure on R^N is Lebesgue measure $\nu(S)$, characterized by the fact that if S is a rectangular solid with sides of lengths λ_i, $i=1,\ldots,N$, then $\nu(S)=\Pi_i\lambda_i$.

If a measure μ on R^N is zero on every set for which ν is zero (i.e., it is absolutely continuous), then the ratio of measures $\mu(H)/\nu(H)$ where H is a neighborhood (with non-zero measure) around a point x is well-defined, and in the limit, as the neighborhood shrinks to measure zero, the ratio f(x), if it exists, is the *density* of μ, also known as the Radon-Nikodym derivative. Not all measures have densities; think of a probability function with a point mass at zero. As H shrinks around the point mass, $\mu(H)$ cannot become less than the point mass, but $\nu(H)$ goes to zero; the density becomes infinite.

Multifractals, as measures, tend to be extremely ill-behaved, not characterizable in terms of densities and point, line, plane, etc., masses.

The simplest way to create a multifractal is by a *multiplicative cascade*. Consider the "binomial multifractal," constructed on a half-open unit interval (0,1] with uniform density as follows: Divide the interval into two halves (open on the left) of equal length. Distribute $0<p<1$ of the mass uniformly on the left half and 1-p of the mass uniformly on the right half (here p is a constant throughout all stages of the construction). Repeat on each subinterval. Figure 2 shows several stages of construction with p=1/3. The horizontal axes show the unit interval and the vertical axes show density. The upper left panel shows stage 1, where 1/3 of the mass is on the left half and 2/3 is on the right. Note that the average density is 1. The upper right panel shows stage 2, where the left and right halves have each been divided. The 2nd and 3rd quarters of the interval have the same density because they have masses of (1/3)*(2/3) and (2/3)*(1/3), respectively. The lower left panel shows stage 4 where the interval has been divided into $2^4 = 16$ subsegments. Some local maxima seem to be appearing, they are labeled. The lower right panel shows stage 7, and begins to give a sense of what the ultimate multifractal looks like. Note the similarity of left and right halves.

As you can see, at the local maxima, the density "blows up" as the scale resolution gets finer. Note how the maximum density increases from panel to panel. However, the rate of divergence is different at different points. The set of locations with particular (different) rates of divergence turn out to be fractals (with different fractal dimensions). Thus we have layers of fractals representing different "orders of singularities," with a relationship between the rate of divergence and the fractal dimension. See Appendix A for mathematical details.

This relationship is known as the *spectrum of singularities* – no single fractal dimension suffices to characterize the fractal measure, hence the name *multifractal.*

Having a spectrum of singularities means that the multifractal measure consists of infinitely spiky peaks sprinkled throughout predominant valleys, but that with proper mathematical technology, the peaks can be classified by the rate at which they diverge to infinity, and comparable peaks can be collected together into fractal "mountain ranges."

Figure 3 shows a real-world density field that approximates a multifractal. It is the population density of the northeastern USA. The big spike in the middle is New York City. Lesser spikes pick out other densely-populated cities.

In their analysis of turbulent meteorological phenomena, [Schertzer & Lovejoy] write the functional relationship between a chosen scale of resolution λ and the average densities φ_λ measured at that scale as:

$$\Pr\{\varphi_\lambda > \lambda^\gamma\} \propto \lambda^{-c(\gamma)} \tag{2}$$

This is very much in the spirit of box-counting for fractals, except the equivalent formulation for fractals would have (1) the event inside Pr{ } being the probability of finding *any* point of the fractal in a λ-neighborhood, instead of points that satisfy a certain degree of singularity, and (2) the exponent on the right hand side being a constant, *the* fractal dimension of the set, instead of a function. In this formulation, the function $c(\gamma)$ carries all the information necessary to characterize, in a statistical sense, the multifractal.[5]

---

[5] It is tempting to read this equation as a statement about the probability of encountering a point with exponent γ or higher or the probability of fractal dimension. However, if the fractal dimension of points having exponent γ or higher is less than the dimension of the embedding space, then such points make up a set of (Lebesgue or probability) measure zero. In the typical multifractal, "almost" all the mass is concentrated in "almost" none of the region. The equation is really a statement about the *scaling relationship* between intensity and probability.

Compare Figure 3 with Figures 4 and 5. The former measures population density at the resolution of 8 miles. The latter two measure it at resolutions of 16 and 32 miles, respectively. Clearly, one's impression of this density field is largely driven by the scale of resolution used. A systematic investigation of the appearance of a field using various scales of resolution is at the heart of multifractal analysis.

A box-counting approach developed by [Lavallée et. al.] known as *Probability Distribution Multiple Scaling* (PDMS) can be used to estimate the probabilities of singularities with assorted rates of divergence. (See also [Lovejoy & Schertzer 1991].) It turns out that directly estimating $c(\gamma)$ in such a fashion is not a productive approach to analyzing real data sets for multifractality due to the severe demands that the procedure places on the sample data. In the next section, we will show how multifractals can be understood equally well through the behavior of their moments.

[Pecknold et. al.] give many examples of (apparent) multifractals in nature. See also [Ladoy et. al.] These include rain and cloud fields (measured from scales of a thousand kilometers and years down to millimeters and seconds – see [Lovejoy & Schertzer 1991]), human population density (as above, also see further discussion below), and foreign exchange rates. Part of the impetus for the development and practical application of multifractal analysis came from "the burgeoning mass of remotely sensed satellite and radar data" [Tessier et. al., 1993]. Depending on the scale of resolution used, measurements of cloud cover could be made to vary drastically; moreover, how this variation with scale behaved was also dependent on the level of intensity chosen as a threshold – just the sort of fractals-within-fractals behavior to be expected from multifractal fields.

Spatial Fields

In this section, we delve into the general theory of self-similar random fields, focusing on the two-dimensional case. (The extension to three or more dimensions is straightforward.) Examples are taken from our applications in property-liability insurance.

Analysis of Multifractal Fields

Analysis of random multifractals is an extension of the analysis of random fields. Recall that a random field $\varphi(\boldsymbol{r})$ is a collection of real-valued random variables φ, indexed by $\boldsymbol{r}$, where $\boldsymbol{r}$ may be

- an integer, for example, in the case where the random field is a (discrete) time series,
- a real number, for example, in the case where the random field is a (continuous) stochastic process,
- a vector in D-dimensional Euclidean space $\mathbf{R}^D$, in the case of a general random field.

Typically, we focus on $D = 1$ for financial/econometric time series and $D = 2$ for spatial distributions in geography or meteorology.

To analyze a random multifractal, we must first respect the fact that it is a measure, and strictly speaking may not (typically does not) possess real-valued densities. Therefore, we cannot treat a random multifractal as a random field $\varphi(\mathbf{r})$.[6] However, as we have seen in previous sections, *when viewed at a finite scale of resolution L*, a multifractal does have a well-behaved density that we can treat as a random field $\varphi_L(\mathbf{r})$. Thus, the approach to studying random multifractals is to consider sequences of random fields that describe the density of the measure at various scales of resolution, and to study the scaling behavior of those sequences.

Appendix B outlines the mathematics. The box-counting approach appears to admit straightforward application (and becomes PDMS) as discussed above. For various reasons discussed below, it is more fruitful to deal with *moments* of the random fields. The key object of the analysis is the so-called *K(q)* function, describing the scaling behavior of the q^{th}-moments of the sequence of random density fields as the scale of resolution λ varies. At finer resolutions, the density fields appear more "spiky" and average q-powers of the fields for q>1 (q<1) get arbitrarily large (small) according to the power law:

$$E(\varphi_\lambda^q) = (\lambda)^{-K(q)}. \tag{3}$$

The boundary conditions $K(0) = K(1) = 0$ further constrain the *K(q)* curve.

Synthesis of Multifractals; 2-D Multiplicative Cascade

Above, we described how recursive application of multiplication of densities – a multiplicative cascade – generated the simple binomial multifractal on a line segment. A similar operation, in two dimensions, can be used to generate a multifractal akin to the Sierpinksi triangle. Consider the following two-by-two matrix:

$$a = \begin{bmatrix} 2.0 & 1.4 \\ 0.6 & 0 \end{bmatrix} \tag{4}$$

Take a unit square with uniform density. Divide it into four quadrants and multiply the density in each quadrant by the corresponding element of a. Note that the average of the four elements of a is 1.0, so the average density across the entire square is unchanged. Repeat the procedure on each quadrant, recursively. In the limit, we have a multifractal. At stage k, neighborhoods of the upper left corner have average density 2^k. That point has the highest degree of singularity.[7] The lower left corner has a different sort of singularity, with density 0.6^k approaching zero as the scale shrinks. The entire lower right half is empty (density zero). Like the Sierpinski triangle, in fact, the square is almost everywhere empty: at each stage, the area with nonzero density is $(3/4)^k$ which approaches zero as k increases without bound. Figure 6 depicts the result.

---

[6] It might be tempting to consider a random measure as a collection of random variables indexed by *subsets* of the underlying $\mathbf{R}^D$ space, but that quickly becomes awkward to work with.

[7] Countably many other points have the same degree of singularity. These are the "upper left corners" of nonzero subcells; at all stages k after some stage a, they have density $m2^{k-a}$.

A random version of the Sierpinski multifractal can be seen in Figures 7 and 8. Here, the positions of the elements of a are randomly shuffled at each downward step in the cascade. Statistically, the random and regular versions are identical, but visually, the random version suggests phenomena taken from biology or geography.

Figure 9 shows the empirically fit and theoretical ("universal") $K(q)$ curves for the Sierpinski multifractal. The latter will be explained in the next section.

Universality Classes; Form of *K(q)*

By making certain plausible assumptions about the mechanisms generating a multifractal, we can arrive at a "universal" theory, akin to a central limit theorem, for multifractals. The critical assumption is that the underlying generator (analogous to the multiplicative factors in the matrix of the previous example) is a random variable with a specific type of distribution: the *exponentiated extremal Lévy* distribution. This is plausible because Lévy distributions generalize the Gaussian distribution in the central limit theorem.

This leads to a two-parameter family of $K(q)$ curves:

$$K(q) = \begin{cases} \dfrac{C_1}{\alpha - 1}(q^{\alpha} - q) & \alpha \neq 1 \\ C_1 \cdot q \log(q) & \alpha = 1 \end{cases} \tag{5}$$

where C_1 acts as a magnification factor and α, related to the tail index of the Lévy generator, determines curvature. These parameters in turn can be related to position and scale parameters μ and σ to be applied to a "standard" Lévy variable $\Lambda_{\alpha}(-1)$.

The derivation, and an introduction to Lévy variables, is presented in Appendix C.

Synthesis of Multifractals: Extremal Lévy Generators

In creating multifractals for liability applications, we adopt this still somewhat controversial theory of universality.[8] That is, each step of a simulated multiplicative cascade is a multiplication by the random factor a given by equation 32 (Appendix C) for appropriately chosen parameters. A cdf of random step factors corresponding to the best universal fit to the Sierpinski cascade example above is shown in Figure 11 (thick curve). A multiplicative cascade with these random step factors could be used instead of the four-element array used above (shown as thin line step function) to construct a multifractal with roughly the same properties as the Sierpinski multifractal.

The Laplace transform of the logarithm of these factors take on the particularly simple forms described in Appendix C. This fact is exploited in data analysis, as will be explained later in the discussion of Trace Moments.

---

[8] The scope and relevance of the necessary conditions to real-world phenomena are hotly debated.

Example Spectrum Analysis: Insured Property Portfolios

A preliminary step, to be taken before fitting a *K(q)* curve to suspected multifractal data, is spectrum (Fourier) analysis. The key point is that a multifractal must possess a spectral density having a certain shape: a straight line in a log-log plot. Furthermore, the slope of that line has additional implications. Therefore, spectrum analysis is used as a screening step before applying multifractal analysis. The mathematics relating *K(q)* to spatial spectral density is presented in Appendix D.

The spatial distribution of the human environment has been studied in geography and human ecology. [Major] analyzed homeowners insurance property as a two-stage Poisson process. Multifractal approaches include the analysis by [Tessier et al. 1994] of the global meteorological network (i.e., locations of weather stations) and [Appleby]'s study of population in the USA and Great Britain. Until [Lantsman et. al.], no one had studied the spatial distribution of insured property values (Total Insured Value, or "TIV").

[Lantsman et. al.] show that some portfolios of insured homeowners properties display a spatial distribution consistent with multifractal behavior (over appropriate scales). Figure 12 shows the isotropic power spectra of the insured value density of five geographically distinct regions of an insured property portfolio.

The preparation of such graphs starts with a grid of insured values at a sufficiently small scale of resolution. First, accumulate insured value totals over a 2^{Tm}-squared grid over the UxU area. In practice, we have found Tm=7 or 8 to be comfortable for Pentium-III class machines. If the data originates as individual observations (e.g., geocoded lat-lon locations) then each observation must be assigned to the appropriate grid cell. If the data originates as areal data (e.g., accumulated values for polygons) then the data must be *allocated* to the grid. In any case, make sure that $L=U/2^{Tm}$ is larger than the resolution of the data. For analysis of large portfolios with ZIP-level data, we typically use U = 512 or 1024 miles[9] with Tm = 6 or 7, resulting in a resolution of L=8 miles, which is a bit bigger than the square root of the average area of a ZIP code.[10]

The second step is to compute the 2-dimensional discrete Fourier transform (DFT) of the array. The third is to convert to an isotropic power spectrum. Appendix D has details. Roughly speaking, the isotropic power spectrum reveals the strength (vertical axis) of various periodicities (horizontal axis) in the spatial data, averaged over all directions.

The horizontal axis of Figure 12 represents the wavenumber (spatial frequency) r where, e.g., wavenumber r =10 corresponds to a periodicity of 512/10 = 51.2 miles. The plots stop at the finest resolution of 8 miles, corresponding to wavenumber r = 512/8 = 64. The vertical axis represents the power (spectral density – i.e. Fourier component amplitude – squared) $P(r)$, with arbitrary constant factors used to separate the five curves.

---

[9] A 1024-mile square covers about one-sixth of the USA.

[10] Since most of the population resides in smaller, more densely populated ZIP codes, we feel that an 8-mile resolution is appropriate.

All but one curve show the smooth, loglinear relationship between power and wave number that is to be expected from a self-similar random field. The exception displays higher than expected spectral amplitude at wavenumbers 45-50 (~11 miles) and less than expected at wave numbers 30-35 (~16 miles). This anomaly was traced to unique factors in this insurer's distribution channel. They had a strong affinity marketing program for military personnel. In Washington DC proper, the portfolio's spatial density of insured value was nearly zero. However, in two suburban enclaves adjacent to nearby military bases, the value density was among the highest observed in the region. The two groups were about 11 miles apart and 16 miles away from the center of DC. If not for this unusual geographic structure to the market, the power spectrum would have been similar to that of the other regions.

Fitting *K(q)*; Trace Moment Analysis

In this section, we discuss how to fit a universal *K(q)* curve to spatial data and use the US population density in the northeast (Figure 3 discussed previously) as an example. Conceptually, the idea is very simple: construct an empirical *K(q)* curve by measuring the moment scaling behavior as expressed in equation 14 (Appendix B), then find parameters C_1 and α that produce a best-fitting theoretical *K(q)* curve (equation 5, equation 31 of Appendix C). In practice, a few wrinkles emerge.

Data preparation starts with the gridding process discussed above in the context of spectral analysis. Most of the square grid should contain meaningful data; too many "structural zeroes" (e.g., representing water or other area that cannot by definition support positive values) will distort the analysis. In the case of Figure 3, each grid entry is an approximate[11] count of persons living in that 8x8 mile geographic square.

Having represented the field as a 2^{Tm}-square matrix, normalize the entries by dividing each by the average value of all the entries; this makes the average entry equal one.

The next step is to prepare a series of locally averaged ("dressed") versions of the data matrix, each 2^T on a side for T=0,2,...,Tm-1. Specifically, the four elements indexed by (2*r+i, 2*c+j) (where i=0,1 and j=0,1) of the 2^{T+1} grid are averaged to become the value of the (r,c) element of the 2^T grid. These represent the same field, but at progressively coarser scales of resolution.[12] See Figures 3 through 5, mentioned previously. Note that for each grid, the average cell value is one. The coarsest grid, corresponding to T=0 and scale U, consists of the single entry, one.

The fourth step is to compute q^{th} powers of the dressed fields and look for a loglinear relationship between them and the scale. If multifractal scaling is present, we should see, for each fixed q, a linear relationship between T (the label identifying the coarseness of a

[11] Recall the original data was at the ZIP code level of resolution, so entire ZIP codes were allocated to particular grid squares, introducing a bit of distortion at the smallest scales.

[12] As a refinement of this process, we start with two grids, the 2^{Tm}-sided grid as described, as well as a slightly coarser $3*2^{Tm-2}$-sided grid, and operate on them in parallel. This way, we get a factor of 1.5 or 1.33 (ideally it would be the square root of two) between adjacent scale ratios instead of a factor of two. This doubles the sample of scale ratios in the analysis.

grid, equal to $\log_2$ of the number of rows or columns in the grid) and the logarithm of the average of the q^{th} power of the grid entries.

Figure 13 shows this relationship for q = 0.6, 0.9, and 1.4. These so-called *trace moments* are close enough to linear to make the multifractal model appropriate.

Having satisfied ourselves that scaling is present, the fifth step is to estimate *K(q)* values as coefficients in a linear regression version of equation 16 (Appendix B), for each of a range of values for q. A certain amount of judgment is called for, however, in choosing the range over which the regression should be carried out. [Essex] and [Lavallée et. al.] discuss "symmetry breaking" that results from the limitations of sample data. The selected range of scaling must avoid these extremes in order to deliver unbiased estimates of moments, and hence undistorted *K(q)* estimates. Linear regression in this case suggests that K(0.6) = -0.2, K(0.9) = -0.1, and K(1.4) = 0.3. An example of the resulting *empirical K(q) curve* based on slopes estimated from regressions of trace moments corresponding to q values ranging from 0.16 to 4.5 is shown in Figure 14.

Before considering how to best fit a universal *K(q)* to the empirical curve, we must address additional limitations of the methodology. The relation between *K(q)* and *c(γ)* (the latter "box counting" exponent expressing the scaling behavior of probability of extreme values) is given by a Legendre transform; there is a one-to-one correspondence between moments and orders of singularities [Tessier et. al. 1993]. Realistic limitations to data (rounding low values to zero, finite sample size, bounded sample) can limit the range of observable singularities and consequently introduce distortions in the measured *K(q)*. In addition, estimating the universal parameters C_1 and α by nonlinear least squares may run afoul of a substantial degree of collinearity between the parameters.

For such instances, [Tessier et. al. 1993, 1994] developed the *double trace moments* technique. This is based on the observation that if a universal field is exponentiated first by η, then averaged to scale λ, then exponentiated to q, we have the relation

$$K(q,\eta) = \eta^{\alpha} \cdot K(q,1) \qquad (6)$$

where the second arguments refer to the exponent of the original field from which the *K(q)* estimate is made. This allows an estimation of the field's α by fixing q and varying η. Figure 15 shows a graph of log K(q,η) vs. log η for various q. Due to the limitations cited above, this equation as applied to sample data tends to break down except for a limited range of η; thus we estimate α as the *maximum* slope observed in the graph. With a good estimate of α in hand, an ordinary least-squares estimate of C_1 is easy to obtain.

In this case, a standard two-parameter nonlinear regression does fine, with α =0.66 and C_1 = 0.72 obtained. The resulting theoretical *K(q)* curve is compared to the empirical version in Figure 16.

Simulating Universal Multifractal Data; Synthetic Geocoding

The utility of a model of insured value emerges when detailed geographical information about a portfolio of risks is lacking. Often the information fed into catastrophe models in the US is based on aggregations at the county or ZIP code level. While this may suffice for hurricane risk analysis, it does not for thunderstorm wind, tornado or hail perils. On the one hand, the average size of a ZIP code is 8 by 8 miles, and the distribution of properties over the area is typically very sparse, irregular and non-uniform. A damage potential (expected damage rate) field representing a hail or tornado event is of a comparable scale (scattered patches less than a mile wide by a few miles long for hail; narrower and longer for tornadoes), and it, too, is highly non-uniform (e.g., 90% of the damage potential from a tornado occurs in less than 5% of its area). Given that the details of the hazard and exposure fields must be superimposed to obtain a reasonable estimate of losses sustained, one can appreciate the difficulty of working with aggregate data.

Previous solutions to the problem were simplistic and reflected a characterization of TIV over the area either as regularly or randomly uniformly distributed, or, at the other extreme, concentrated at a single point, (i.e., the area's centroid). The result of this kind of misrepresentation is a critical misestimation of the variability inherent in the process of loss generation. Figures 17 and 18 illustrate this.

Figure 17 is a map of a portion of a real homeowner property portfolio. The scale is 8 miles on a side, the average size of a ZIP code. Figure 18 shows a realization of the same number of homes assuming a uniform spatial point process. The true portfolio shows more "clumps and gaps" than the relatively smoother uniform random version. Figure 19 shows the results of applying the multifractal model. While it does not reproduce the original portfolio (no random model would be expected to), it does appear to exhibit the same spatial statistics. When intersected with a number of simulated damage footprints from hail or tornadoes, it will clearly do a better job of estimating the damage probability distribution than will either the uniform random version or a version that puts all the properties at the center of the figure. The uniform distribution will result in too many small loss events and not enough large loss events, and vice-versa for the centroid.

The construction of a synthetic geocoding proceeds as follows:

1. Create a multifractal field over the area in question. Typically, we use a five- to seven-stage process, depending on the outer scale. A seven-stage process divides a square into $2^7 x 2^7$=16,384 grid cells; this is sufficient to carve an 8-mile square into 2.5 acre parcels. At each stage i = 0 to T_m, instantiate a $2^i x 2^i$ array of independent and identically distributed exponentiated extremal Lévy random variables (see equation 32 of Appendix C).[13] In the example of Figure 19 we used the parameters $\alpha = 0.8$, $C_1 = 0.6$. In [Lantsman et. al.], we reported different parameters for industry and selected client portfolios.[14] Combine factors via multiplicative cascade as described for the Sierpinski multifractal.

[13] [Samorodnitsky & Takku] has an efficient algorithm for simulating Lévy variables.

[14] Specifically, $\alpha = 1.024$ and $C_1 = 0.560$ for industry TIV measured at the ZIP code level, and $\alpha = 0.552$, $C_1 = 0.926$ for a geocoded client portfolio. The implications of this difference are discussed in that paper.

2. Normalize the field and use it as a probability distribution to drive a multinomial point process. If the area is a polygon other than a square, then grid cells must be identified as to being inside or outside the polygon. Outside grid cells are zeroed out; inside cell intensity values are divided by the total of all inside values to renormalize. Say the grid probability in cell i is p_i. The desired number of homes, N, is then allocated to each cell N_i, by a multinomial($N, p_1, p_2, \ldots, p_{4^{\wedge}Tm}$) joint random variable draw. In practice, this is implemented by a sequence of conditional binomial r.v. realizations. The first r.v. is N_1~binomial(N, p_1). Subsequent cells' realizations are conditional on all that precede, viz., N_3~binomial($N-N_1-N_2, p_3/(1-p_1-p_2)$), etc.

Project APOTH: Thunderstorm Simulation

Occurrence rates for small scale thunderstorm-related perils (wind, hail, tornadoes) have traditionally been computed as an annual rate averaged over a fairly wide region. This is done by counting the number of occurrences of some peril of interest – say, hail two inches or more in diameter, tornadoes F3 or more, etc. – in a given area (frequently, a one- or two-degree longitude/latitude box). This count is normalized to the area of the box and the period of record. When this process is continued for more boxes (usually overlapping), contour maps can be plotted showing the smoothed variation in the rate. These types of maps are often developed for differing severity levels, such as hail >1", >2", etc. or tornadoes >F2, >F3, etc. To this extent, both frequency and severity are incorporated into them.

Maps such as these are often used to estimate the probability of occurrence for an event of at least a certain severity at a single location. Such an application might be estimating the chance that a nuclear power plant will be hit by an F4 or F5 tornado. These maps can also be used to estimate probabilities of an event hit to a town or subdivision.

Unfortunately, point-frequency maps are not very useful for modeling the typical insurance catastrophe loss event. While there are cases where a single violent tornado or a single storm of large diameter hail hits a highly populated area and produces a large loss, there are also other cat events whose losses are aggregates of many moderate losses which occurred in different locations, possibly over several states and over several days.

The goal of Project APOTH (Atmospheric Perils Other Than Hurricane) is to develop the capability to credibly estimate probabilistic losses from the thunderstorm perils of hail, tornado, and straight-line winds (non-tornadic high wind gusts). The APOTH project presently has models that can realistically simulate both the geographical and seasonal characteristics of severe storms, as well as model their small-scale ground damage patterns as they affect homeowners anywhere in the lower 48 states of the USA. The model can be easily extended to include other lines of business once vulnerability functions become available from further research.

One objective of natural hazard simulation is to produce a "future history" of meteorological events, in sufficiently rich detail to be able to explore the range of damage effects on a subject portfolio of insured properties, yet maintain a statistically stationary relationship to the available historical data upon which the simulation is based.

The production of tornadoes and hail involves meteorological processes exhibiting complex behavior over a wide range of scales, from synoptic weather patterns (thousands of km) down to the size of the hailstone (millimeters or centimeters). We have made use of multifractal modeling, not only to distribute property values in statistically appropriate patterns, but directly in the simulation of the hazards themselves.

Multifractal modeling is not appropriate to all scales, however. Thunderstorms exhibit a strong seasonality during the year, nonhomogeneity of occurrence frequencies over distances of thousands of km, and anisotropy in terms of preferred directions of movement. At the smaller scales, the structure of tornado tracks and hail streaks (continuous bands of hailfall) are also highly idiosyncratic. In between, however, we have found that the scale of the *swath* (tens to hundreds of km) on a single day is amenable to multifractal modeling.

Figure 20 shows a set of reported hail occurrences for 3/30/98. Unfortunately, while swaths may make conceptual and meteorological sense, data are not reported in swath groupings. Before we can analyze swaths, we must identify them, using various tools including Bayesian classification, modal clustering, and nearest-neighbor methods. Figure 21 shows the same set of reports, now grouped into meaningful swaths.

In order to expand the data into a meaningful set of possible alternative scenarios, we have followed the practice of other modelers in using the historical data as a template for a synthetic "probabilistic database" of possible events. Figure 22 exemplifies the typical practice of equally displacing all reported events by some random X-Y translation vector.[15] One of our innovations is to use multifractal modeling to create and recreate alternative detailed patterns within a given swath.

Our procedure is as follows:

1. Historical reports are grouped into swaths as mentioned above.
2. Swaths are characterized by a small number of key parameters: the location, size, orientation, and eccentricity of the bounding ellipse; the prevailing storm motion direction within, and parameters describing the overall intensity level of the activity. In the case of hail, intensity is defined by a categorical type label and the total volume of hail (number of hailstones). In the case of tornado, intensity is defined by Fujita class-specific Poisson parameters for the number of touchdowns and two principal component scores defining the conditional distribution of tornado path lengths. In the case of non-tornado wind, intensity is defined as total wind power (watts).
3. When an historical swath is drawn from the database as a template for a simulated swath, the ellipse is gridded at the 1-km scale and a multifractal field (with parameters appropriate to the peril and type) is laid down over the grid. As described above for simulated geocoding, this field is "condensed" to a schedule of report (hail, tornado, or wind event) locations.

[15] Since this translation is by no more than a degree in either direction, it is a bit difficult to see at first.

4. Details of each report (hail streak size and intensity details; tornado F-class and track length, etc.) are drawn from conditional distributions, with correlation induced with the intensity of the underlying multifractal field at the point of condensation.

Figure 23 shows several realizations of the multifractal simulation of these particular swaths. Note how they respect the ellipse boundaries, yet vary dramatically in their inner detail. A much richer variety of possible outcomes is made possible, compared to simple location-shift models, but the statistics of event properties and their spatial colocation are still respected.

Conclusion

In this part I paper, we introduced the ideas of fractal point sets and multifractal fields. We showed that while those mathematical constructs are rather bizarre from a traditional point of view (e.g., theory of smooth, differentiable functions), they nonetheless have applicability to a wide range of natural phenomena, many of which are of considerable interest to the casualty actuary. We showed how to analyze sample data from multidimensional random fields, detect scaling through the use of the power spectrum, detect and measure multifractal behavior by the trace moments and double trace moments techniques, fit a "universal" model to the trace moments function $K(q)$, and use that model to simulate independent realizations from the underlying process by a multiplicative cascade. In particular, we discussed synthetic geocoding and the simulation of non-hurricane atmospheric perils.

In the companion part II paper, we focus on time series analysis and financial applications.

Appendix A: Binomial Multifractal

This appendix establishes a relationship between orders of singularities and fractal dimension in the binomial multifractal on the half-open unit interval (0,1]. We follow the presentation in [Mandlebrot 1989].

Divide the interval into two halves (each open on the left) of equal length. Distribute 0<p<1 of the mass uniformly on the left half and 1-p of the mass uniformly on the right half (here p is a constant throughout all stages of the construction). Repeat on each subinterval.

At stage k of the construction, we have 2^k pieces of length 2^{-k}, of which k!/(h!(k-h)!) of them have density $p^h(1-p)^{k-h}$.

Any point x in the interval can be expanded as a binary number $0.b_1b_2b_3\ldots$[16] By considering the sequence of expansions truncated at b_k, we make meaningful statements about the behavior of the measure at x. For example, define

$$f(k) \equiv \frac{1}{k}\sum_{i=0}^{k} b_i \,. \tag{7}$$

Then, in a $\lambda=2^{-k}$ -wide neighborhood of x, the average density is $p^{1-f(k)}(1-p)^{f(k)} = \lambda^{-\alpha(k)}$, where $\alpha(k) = \log_2(p^{1-f(k)}(1-p)^{f(k)})$. If $f \equiv \lim_{k\to\infty} f(k)$, the proportion of 1's in the binary expansion, exists, then we can consider that the density behaves as $\lambda^{-\alpha}$ in the limit. Such a point is termed a *singularity of exponent α*. As λ gets smaller, the density may grow without limit or shrink to zero, but the rate of that growth is controlled by the exponent α, a quantification of spikiness at that point. (This is the classical Hölder exponent.)

What is the fractal dimension of the set of such points? (Here, the exposition becomes quite deliberately sketchy, as proper delta-epsilon arguments would take up an undue amount of space.) At stage k, we have a total of 2^k intervals, of which n = k!/((kf)!(k(1-f))!) have density defined by $f(k) = f$. Recalling Stirling's approximation for factorial,

$$x! \approx \sqrt{2\pi}\cdot\exp(-x)\cdot x^{x+\frac{1}{2}} \tag{8}$$

we can write an approximation for n as

$$n(\lambda) = \frac{\left(f^f(1-f)^{1-f}\right)^{-k}}{\sqrt{2\pi k f(1-f)}} \tag{9}$$

This gives us a fractal dimension d = f $\log_2 f$ + (1-f)$\log_2$(1-f). Since the exponent α = f $\log_2$(1-p) + (1-f) $\log_2$(p) is also a function of f, we have a functional relationship between the order of the singularity α and the fractal dimension d of the set of points having that exponent.

[16] Since binary xyz0111... is the same as xyz1000..., let us agree to use only the 111... representation for such cases. (This is consistent with our closing the right side of intervals.)

Appendix B: Analysis of Multifractal Fields

A random field is called *stationary*[17] if the distribution of $\varphi(\boldsymbol{r}_1)$ is the same as that of $\varphi(\boldsymbol{r}_2)$ for any different $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. This does not imply the two random variables are independent, however. For example, a multivariate normal may have identical marginal distributions but nonetheless possess a nontrivial correlation structure. A nonstationary field is said to have *stationary increments* if the distribution of $\varphi(\boldsymbol{r}_1) - \varphi(\boldsymbol{r}_2)$ depends only on the difference vector $\boldsymbol{r}_1 - \boldsymbol{r}_2$. Furthermore, for $D > 1$, such a field is said to be *isotropic* if the distribution of $\varphi(\boldsymbol{r}_1) - \varphi(\boldsymbol{r}_2)$ depends only on the *magnitude* of that vector, $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$.

Our discussion follows [Novikov & Stewart], [Shertzer & Lovejoy], [Marshak et. al.] and [Menabde et. al.] in the general context of a D-dimensional space and for stationary fields. The generalization to non-stationary fields will be discussed in Appendix D.

Formally, consider a measure $\mu(X)$ whose domain consists of a σ-field of subsets X of $\mathbf{R}^D$. Define the scale-L average density as:

$$\varphi_L(r) = L^{-D}\mu(V) \tag{10}$$

where V is a D-dimensional hypercube of side length L centered at $\boldsymbol{r}$. Our first condition for μ to be a random multifractal is to assume $\varphi_L(\boldsymbol{r})$ is a random field. For a particular realization of μ, we can compare such field realizations at different scales of resolution L and U by considering their relative densities[18] defined as:

$$a_{L,U}(r) = \varphi_L(r)/\varphi_U(r) \tag{11}$$

where $L < U$, therefore $V_L \subset V_U$. This is only defined for nonzero values of φ_U, but note that when it is zero, so must be φ_L. We have the property that:

$$a_{L,U} = a_{L,\rho}a_{\rho,U} \tag{12}$$

where $L < \rho < U$ (therefore $V_L \subset V_\rho \subset V_U$), and we have suppressed mention of volume centers $\boldsymbol{r}$. This is true for any realization, and therefore can be considered a statement about the random variables as well.[19]

These relative densities are random fields in their own right. They characterize how the fluctuation (or *intermittency*) of the field varies as a function of scale. The assumption of

---

[17] Physicists would use the word *conservative*.

[18] These would be known as Radon-Nikodym derivatives to a statistician or *breakdown coefficients* to a physicist.

[19] It is helpful to think of the measure μ as a physical quantity, such as mass, rather than a probability measure. That way, probability statements about the random μ will not be confused with statements about the properties of particular realizations of μ.

stationarity implies that $a_{L,U}$ is a random variable whose distribution does not depend on the position of the volume center $\boldsymbol{r}$. Furthermore, we assume it depends only on the *ratio* L/U and that the random variables $a_{L,\rho}$ and $a_{\rho,U}$ in equation (3) are independent. This last statement is the technical definition of $\mu(X)$ being a *statistically self-similar* (a.k.a. *scale-invariant*, or *scaling*) random measure.

Scaling of Moments, *K(q)* Function

It is possible to show that under these assumptions the statistical moments of $a_{L,U}$ have the property:

$$E(a^q_{L,U}) = E(a^q_{L,\rho})E(a^q_{\rho,U}) \tag{13}$$

where $E(\cdot)$ is expected value operator. Since the moments of $a_{L,U}$ depend only on the ratio L/U, the most general expression for scaling behavior of statistical moments is:

$$E(a^q_{L,U}) = (L/U)^{-K(q)} \tag{14}$$

with the boundary conditions $K(0) = K(1) = 0$.

It is worth noting that for some processes $K(q) = \theta \cdot (q-1)$ (for $q>0$) which is usually referred to as *simple scaling* or the *(mono)fractal* case. However, in nature, most processes exhibit a more complex behavior and the $K(q)$ function evaluation requires a more elaborate approach. There is a wide class of random multiplicative cascade models that produce *multiscaling* behavior and *multifractal* fields [Parisi & Frisch].

For the special case where scale steps are factors of two, we can specialize equation 14.

From the definition of $a_{L,U}$ in equation 11, and noting that φ_U is equal to one, we can write:

$$E(a^q_{U\cdot 2^{-T},U}) = E(\varphi^q_{U\cdot 2^{-T}}) = 2^{T\cdot K(q)} \tag{15}$$

or

$$\log_2\left[E(\varphi^q_{U\cdot 2^{-T}})\right] = T \cdot K(q) \tag{16}$$

This form reveals $K(q)$ as the coefficient in a log-linear regression between the scale index T and average q-power of the field, as used in empirical data analysis.

Appendix C: Universality Classes; Form of *K(q)*

To further explore the structure and behavior of the *K(q)* function we follow [Schertzer & Lovejoy], [Lovejoy & Schertzer 1990], and especially [Menabde et. al.] and formalize the idea of a multiplicative cascade generator (MCG):

$$G_{L,U} = -\ln(\varphi_L L^D / \varphi_U U^D) \tag{17}$$

We assume that the measure is not zero on any finite hypercube, therefore G is everywhere defined. By definition of φ_L (equation 10 of appendix B), the ratio is less than one and so $G_{L,U}$ is a non-negative random variable whose distribution depends only on the ratio *L/U*. For arbitrary *n*, we can introduce *n* hypercubes of side length ρ_n which nest between V_L and V_U so that:

$$L/\rho_1 = \rho_1/\rho_2 = ... = \rho_n/U = (L/U)^{1/n} \tag{18}$$

After a series of transformations the resulting expression for $G_{L,U}$ will be:

$$G_{L,U} = G_{L,\rho_1} + G_{\rho_1,\rho_2} + ... + G_{\rho_n,U} \tag{19}$$

The random variables on the right-hand side of equation (8) are assumed to be independent and *identically distributed* random variables with a pdf:

$$p(g;(L/U)^{1/n}) = p(g;\rho_i/\rho_{i+1}) \tag{20}$$

which depends solely on the scale ratio, $(L/U)^{1/n}$. The property expressed in equations 19 and 20 implies that the probability density for $G_{L,U}$ belongs to the class of *infinitely divisible* distributions [Feller]. The natural candidate for a MCG would therefore be a random variable with a stable Lévy distribution.

An aside on Lévy random variables is in order. Lévy random variables generalize Gaussian (normal) random variables in the Central Limit Theorem. The CLT states that the distribution of a sum of a set of N independent, identically distributed random variables *with finite variance* converges to a normal distribution as the number N increases without bound. More generally, if the restriction to finite variance is removed, we can say that the sum converges to a Lévy distribution.

Lévy distributions are characterized by four parameters: α, which must be in (0,2]; β, which must be in [-1,1]; and μ and $\sigma>0$, which are otherwise unrestricted. The latter two are location and scale parameters, respectively, allowing us to express a Lévy random variable as $\mu+\sigma\Lambda_\alpha(\beta)$ where Λ is "standardized" and depends on only two parameters. Note that σ is not the standard deviation because in general, variance is infinite for a Lévy random variable. The parameter α is the tail index: the case $\alpha=1$ gives the Cauchy distribution while the case $\alpha=2$ gives the Normal distribution. As x increases without bound, the probability that a Lévy random variable exceeds x is proportional to $x^{-\alpha}$. The second parameter, β, is a symmetry index: if $\beta=0$, then the distribution is symmetric; otherwise, the probability of the upper tail is proportional to $1+\beta$ and the probability of the lower tail is proportional to 1-β (in the large-x limit). When α=1, the β parameter becomes irrelevant, and is conventionally set to 0. While there is no closed-form expression for the distribution function for Lévy variables, the characteristic function is analytically tractable.[20]

To develop a moment scaling relation for the random multifractal μ(X) we apply the Laplace transform to the density function $p(g; L/U)$:

$$\psi(s;L/U) = \int_0^\infty \exp(-s \cdot g)\, p(g;L/U)\, dg \tag{21}$$

where $s \geq 0$.

Because $p(g; L/U)$ is the pdf of an infinitely divisible distribution, from equation 18 we can conclude:

$$\psi(s;L/U) = \psi^n(s;(L/U)^{1/n}) \tag{22}$$

Equation 22 has the solution:

$$\psi(s;L/U) = (L/U)^{\chi(s)} \tag{23}$$

where, according to the general properties of infinitely divisible distributions [Feller], $\chi(s)$ can in the most general case be represented by a Lebesgue integral:

$$\chi(s) = \int_0^\infty \frac{1-\exp(-s \cdot x)}{x} M(dx) \tag{24}$$

where M is a measure such that the integral:

$$\int_1^\infty x^{-1} M(dx) < \infty \tag{25}$$

For processes under consideration with some degree of rigor we can limit ourselves to considering only measures M having a density M^*. In such cases we can replace the Lebesgue integral with a Riemann integral, replacing $M(dx)$ with $M^* dx$. It is this density function M^* (or equivalently $\chi(s)$ or $p(g;L/U)$) that completely determines the properties of the MCG and therefore the (statistical) properties of the self-similar multifractal μ(X).

The expression in equation 21 could be considered as an expectation of $exp(-sG_{L,U})$ and can be rewritten as follows:

---

[20] Refer to [Samorodnitsky & Takku] for information on simulating and evaluating Lévy random variables.

$$\psi(s; L/U) = E\left[\exp(-sG_{L,U})\right]$$
$$= E\left[\exp(s\ln(\varphi_L L^D / \varphi_U U^D)\right] = E(\varphi_L / \varphi_U)^s (L/U)^{sD} \quad (26)$$

From equations 23 and 26 we can find the moment scaling relation:

$$E(\varphi_L / \varphi_U)^s = (L/U)^{-sD+\chi(s)} \quad (27)$$

From equations 14 (appendix B) and 27, after replacing s with q, we can get following expression:

$$K(q) = qD - \chi(q) \quad (28)$$

Since by definition in equation 14, $K(1)=0$, one has the normalization condition in equation 28 that $\chi(1)=D$. The asymptotic behavior of $K(q)$ could be deduced from equations 24 and 28 as:

$$K(q) = qD + O(1) \quad (29)$$

One can choose any form for the density measure M^* that satisfies the convergence and normalization conditions of equations 25 and 28. The most appealing measure is:

$$M^*(x) \propto x^{-\alpha} \quad (30)$$

(specifying only the limiting behavior for large x) which corresponds to a stable Lévy distribution [Feller]. With this choice of measure and proper renormalization we can express $K(q)$ as:

$$K(q) = \begin{cases} \dfrac{C_1}{\alpha - 1}(q^\alpha - q) & \alpha \neq 1 \\ C_1 \cdot q \log(q) & \alpha = 1 \end{cases} \quad (31)$$

This expression represents the classes of "universal generators" [Schertzer & Lovejoy]. The first remarkable thing to notice is that a universal generator is characterized by only two fundamental parameters (C_1, α). The idea behind the introduction of universality classes is that whatever generator actually underlies the multiplicative cascade giving rise to a random multifractal, it may "converge" (in some sense) to a well-defined universal generator.

With only two degrees of freedom, the $K(q)$ curves represented by universal multifractals are of a limited variety. As mentioned previously, $K(q)$ is constrained to go through the points (0,0) and (1,0) with negative values when $0<q<1$ and positive values for $q>1$. The parameter C_1 clearly behaves as a vertical scaling factor. The α parameter affects the curvature, as can be seen in Figure 10, with the extreme case of $\alpha \to 0$ converging to a straight line (with discontinuity at $q = 0$).

For this "universality" result to be useful, we must also investigate which classes of MCG are stable and attractive under addition and will at least converge for some positive moments (not necessarily integer order). The task to specify universality classes could be accomplished by considering the Lévy distribution in a Fourier framework, i.e., its characteristic function. The restriction imposed by the Laplace transform (equation 21) is that we require a steeper than algebraic fall-off of the probability distribution for positive order moments, hence, with the exception of the Gaussian case ($\alpha = 2$), we have to employ strongly asymmetric, "extremal" Lévy laws (β = -1), as emphasized by [Schertzer & Lovejoy]. The Lévy location parameter μ is fixed by the normalization constraint and the scale parameter σ is derived from C_1 [Samorodnitsky & Takku]. Roughly speaking, the universality theory states that multifractals built from random multiplicative cascades are statistically equivalent to those built from a special class: the *exponentiated extremal Lévy variables*:

$$a = \exp(\mu + \sigma \cdot \Lambda_\alpha(-1)) \tag{32}$$

According to [Schertzer & Lovejoy], we can designate the following main universality classes by specifying the parameter α:

1. $\alpha = 2$: the Gaussian generator is almost everywhere (almost surely) continuous. The resulting field is a realization of the log-normal multiplicative cascade introduced by [Kolmogorov], [Obukhov], and [Mandelbrot 1972] to account for the effects of inhomogeneity in three-dimensional turbulent flows (turbulent cascades).
2. $2 > \alpha > 0$: the Lévy generator is almost everywhere (almost surely) discontinuous and is extremely asymmetric.
3. $\alpha = 0+$: this limiting case corresponds to divergence of every statistical moment of the generator and represents the so-called "β" model.

Appendix D: Spectrum Analysis; $K(q)$ and Spectrum Slope

In this section, we explore the relation between the moment scaling function $K(q)$ and the power spectrum of the stationary field φ_L that represents a random multifractal at the (sufficiently small) scale of resolution L.[21] Recall that the power spectrum of a time series or one-dimensional stochastic process quantifies the magnitude (amplitude) of cycles of various lengths (periods). Spectral analysis generalizes to multidimensional fields by characterizing not only the amplitude and periodicity of such "waves" but their directions as well. An *isotropic* power spectrum averages the D-dimensional power spectrum over all directions, converting it to a one-dimensional spectrum.[22]

Because of Fourier duality between the correlation function of the field and its power spectrum [Feller] it is customary in analysis of empirical stochastic processes to examine the correlation structure of a process and then map it into Fourier space. But the correlation function is not well suited to analyzing non-stationary fields so we need to develop some guidance as to how to check for stationarity, and, if it exists, how to quantify the underlying field.

Because in the case of stationarity the functional form of the correlation function closely relates to the $K(q)$ function, we can be reasonably confident in establishing a direct link between the power spectrum and $K(q)$ function of the field. Following [Menabde et al.], we demonstrate how it could be accomplished.

For a D-dimensional isotropic random field $\varphi_L(r)$:

$$E(\varphi_L(r_1)\varphi_L(r_2)) = C(|r_1 - r_2|) \tag{33}$$

where $C(r)$ is the correlation function of the field.

The Fourier transform of $\varphi_L(r)$ field is defined as:

$$\psi(k) = \int \exp(-ikr)\varphi_L(r)d^D r \tag{34}$$

where $i = \sqrt{-1}$, the unit imaginary number. For an isotropic field (equation 33) one has that:

$$E(\psi(k)\psi(h)) \sim \delta(k-h)P(k) \tag{35}$$

---

[21] Historically, power spectrum analysis played a central role in identifying and characterizing the scaling properties of self-similar random fields. Recent advances [Marshak et al.] in understanding the limitations of applicability and sensitivity of power spectrum analysis leads one to realize that the issue of stationarity is critical in qualifying and quantifying intermittency of the field. The erroneous assumption that everything could be extracted from knowledge of the spectral exponent leads to a failure to discriminate between qualitatively different fields.

[22] This is explained more fully below.

where $\delta()$ is a delta function (1 at 0, 0 elsewhere) and $P(k)$ is the isotropic power spectrum. On the other hand, from equations 33 and 34 we can get the expression:

$$E(\psi(k)\psi(h)) = \int\int \exp(-ikr_1 - ihr_2)C(|r_1 - r_2|)d^D r_1 d^D r_2 \tag{36}$$

After some mathematical manipulations with integrals involving change of variables, introduction of polar coordinates, and performing the integration over the angular variables, one can obtain the following elegant result for the power spectrum of a stationary isotropic random field:

$$P(k) \propto k^{-D+K(2)} \tag{37}$$

The practical implementation of the above on an NxN square grid V(m,n) of intensity values is as follows: Compute the array:

$$H(k,h) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} \exp\left\{-2\cdot\pi\cdot i\cdot\left(\frac{m\cdot k + n\cdot h}{N}\right)\cdot V(m,n)\right\}. \tag{38}$$

Convert this to the isotropic power spectrum by accumulating values $|H(k,h)|^2$ (i.e. complex magnitude squared) into one-dimensional array cells A(r) where

$$r = round\left(\left|\begin{matrix}(k+N/2)\bmod N - N/2\\(h+N/2)\bmod N - N/2\end{matrix}\right|\right). \tag{39}$$

(Here, the vertical bars indicate vector magnitude, i.e., square root of sum of squares.) Then convert A values to averages P by dividing each accumulated A entry by the number of H cells contributing to the entry.

Equation 37 could be utilized in many ways: to check a D-dimensional stationary isotropic field for SS properties, to verify the validity of a numerical approximation of the $K(q)$ function at the point q = 2, or to examine a non-stationary field with stationary increments (Brownian motion and "fractional Brownian motion"). Note that $P(k)$ and $K(2)$ can be computed by independent methods from the same data, enabling one to verify the consistency of assumptions about stationary increments.

If we relax the assumption of stationarity, the problem of identification and characterization of SS fields develops some complications. We outline some important guidelines in handling non-stationary fields:

1. First of all, the power spectrum analysis still can indicate self-similarity of the field under investigation, revealing the following form:

$$P(k) \propto k^{-\beta} \tag{40}$$

2. For D-dimensional fields the condition $\beta > D$ will indicate lack of stationarity, but some transformations of the original field (like power-law filtering or taking the absolute value of small-scale gradients) could produce a stationary field.

3. The spectral exponent β contains information about the degree of stationarity of the field. The introduction of a new parameter H (sometimes called the Hurst exponent) related to β could aid in the task of characterizing the degree of persistence or long-term memory of the field. We will illustrate the importance of parameter H for time series in the part II paper.
4. The arguments that the correlation function is not well suited for non-stationary situations (because of its translation dependence) led to the development of new ideas about the statistical properties of non-stationary fields to be properly estimated by spatial averaging procedures. The Wiener-Khinchine relation applicable to fields with stationary increments [Monin & Yaglom] states that it is the second-order structure function – not the correlation function – that is in Fourier duality with the power spectrum. We will introduce the structure function in the context of time series analysis in the part II paper and illustrate how the structure function is the one-dimensional analog of the $K(q)$ function.

A further refinement of the multiplicative cascade is to pass from the *discrete* cascade, which is what has been described up to this point, to the *continuous* cascade. The idea behind a continuous cascade is that rather than proceeding in identifiable steps, the multiplicative transfer of intensity variation between scales happens continuously at all scales. [Schertzer & Lovejoy] describe a method of implementing continuous cascades by means of the Fourier transform.

The functional form for $K(q)$ (equation 31 in appendix C) could be extended to nonstationary fields, and fractional integration (power-law filtering in Fourier space) could be used to transform simulated stationary random fields to any desired degree of non-stationarity (in the sense of spectral exponent β). This is considered more fully in the part II paper.

References

Appleby, Stephen. (1996). "Multifractal Characterization of the Distribution Pattern of the Human Population," Geographical Analysis, vol. 28, no. 2, pp. 147-160.

Barnsley, Michael (1988). *Fractals Everywhere,* Boston: Academic Press.

Davis, Anthony, Alexander Marshak, Warren Wiscombe, and Robert Calahan. (1994). "Multifractal Characterizations of Nonstationarity and Intermittency in Geophysical Fields: Observed, Retrieved, or Simulated," *Journal of Geophysical Research* v. 99, no. D4, pp. 8055-8072.

Essex, Christopher. (1991). "Correlation Dimension and Data Sample Size," in *Non-Linear Variability in Geophysics: Scaling and Fractals,* Daniel Schertzer and Shaun Lovejoy, eds., The Netherlands: Kluwer Academic Publishers.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, volume 2,* New York: Wiley.

Kolmogorov, A. N. (1962). "A Refinement of Previous Hypothesis Concerning the Local Structure of Turbulence in Viscous Incompressible Fluid at High Reynolds Number," *Journal of Fluid Mechanics,* vol. 13, pp. 82-85.

Ladoy, P., S. Lovejoy, and D. Schertzer. (1991). "Extreme Variability of Climatological Data: Scaling and Intermittency," in *Non-Linear Variability in Geophysics: Scaling and Fractals,* Daniel Schertzer and Shaun Lovejoy, eds., The Netherlands: Kluwer Academic Publishers.

Lantsman, Yakov, John A. Major, and John J. Mangano. (1999). "On the Multifractal Distribution of Insured Property," to appear in *Fractals.*

Lavallée, D., D. Schertzer, and S. Lovejoy. (1991). "On the Determination of the Codimension Function," in *Non-Linear Variability in Geophysics: Scaling and Fractals,* Daniel Schertzer and Shaun Lovejoy, eds., The Netherlands: Kluwer Academic Publishers.

Lovejoy, Shaun, and Daniel Schertzer. (1990). "Multifractals, Universality Classes and Satellite and Radar Measurements of Cloud and Rain Fields," *Journal of Geophysical Research* v. 95, no. D3, pp. 2021-2034.

Lovejoy, Shaun, and Daniel Schertzer. (1991). "Multifractal Analysis Techniques and the Rain and Cloud Fields from 10^{-3} to 10^{6} m," in *Non-Linear Variability in Geophysics: Scaling and Fractals,* Daniel Schertzer and Shaun Lovejoy, eds., The Netherlands: Kluwer Academic Publishers.

Major, John A. (1999). "Index Hedge Performance: Insurer Market Penetration and Basis Risk," in *The Financing of Catastrophe Risk,* K. Froot, ed., National Bureau of Economic Research, Chicago: University of Chicago Press, pp. 391-426.

Mandlebrot, B. (1972). "Statistical Models of Turbulence," in *Lecture Notes in Physics,* vol. 12, M. Rosenblatt and C. Van Atta, eds., Springer Verlag, p. 333.

Mandlebrot, B. (1982). *The Fractal Geometry of Nature*, Freeman.

Mandlebrot, B. (1988). "An Introduction to Multifractal Distribution Functions," in *Fluctuations and Pattern Formation*, H. E. Stanley and N. Ostrowsky, eds., Kluwer.

Mandlebrot, Benoit B. (1989). "The Principles of Multifractal Measures," in *The Fractal Approach to Heterogeneous Chemistry*, D. Avnir, ed., Chichester: John Wiley & Sons.

Marshak, Alexander, Anthony Davis, Warren Wiscombe, and Robert Calahan. (1997). "Scale Invariance in Liquid Water Distributions in Marine Stratocumulus. Part II: Multifractal Properties and Intermittency Issues," Journal of the Atmospheric Sciences, vol. 54, June, pp 1423-1444.

Menabde, Merab, Alan Seed, Daniel Harris, and Geoff Austin. (1997). "Self-Similar Random Fields and Rainfall Simulation," *Journal of Geophysical Research* v. 102, no. D12, pp. 13,509-13,515.

Monin, A. S., and A. M. Yaglom. (1975). *Statistical Fluid Mechanics, volume 2*, Boston: MIT Press.

Novikov, E. A., and R. Stewart. (1964). "Intermittency of Turbulence and Spectrum of Fluctuations in Energy-Dissipation," *Izv. Akad. Nauk SSSR, Ser. Geofiz*, vol. 3, pp. 408-412.

Obukhov, A. (1962). "Some Specific Features of Atmospheric Turbulence," *Journal of Geophysical Research* v. 67, pp. 3011-3014.

Parisi, G., and U. Frisch. (1985). "A Multifractal Model of Intermittency," in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, Ghil, Benzi, and Parisi, eds., North-Holland, pp. 84-88.

Pecknold, S., S. Lovejoy, D. Schertzer, C. Hooge, and J. F. Malouin. (1998). "The Simulation of Universal Multifractals," in *Cellular Automata: Prospects in Astrophysical Applications*, J. M. Perdang and A. Lejeune, eds., World Scientific.

Samorodnitsky, Gennady, and Murad S. Taqqu, (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, New York: Chapman and Hall.

Schertzer, Daniel, and Shaun Lovejoy. (1991). "Nonlinear Geodynamical Variability: Multiple Singularities, Universality and Observables," in *Non-Linear Variability in Geophysics: Scaling and Fractals*, Daniel Schertzer and Shaun Lovejoy, eds., The Netherlands: Kluwer Academic Publishers.

Tessier, Y., S. Lovejoy, and D. Schertzer. (1993). "Universal Multifractals: Theory and Observations for Rain and Clouds," *Journal of Applied Meteorology*, vol. 32, February, pp. 223-250.

Tessier, Y., S. Lovejoy, and D. Schertzer. (1994). "Multifractal Analysis and Simulation of the Global Meteorological Network," *Journal of Applied Meteorology*, vol. 33, December, pp. 1572-1586.

Wilson, J., D. Schertzer, and S. Lovejoy. (1991). "Continuous Multiplicative Cascade Models of Rain and Clouds," in *Non-Linear Variability in Geophysics: Scaling and*

Fractals, Daniel Schertzer and Shaun Lovejoy, eds., The Netherlands: Kluwer Academic Publishers.

Woo, Gordon (1999). *The Mathematics of Natural Catastrophes,* London: Imperial College Press.

Figures for Part I

John A. Major
Yakov Lantsman

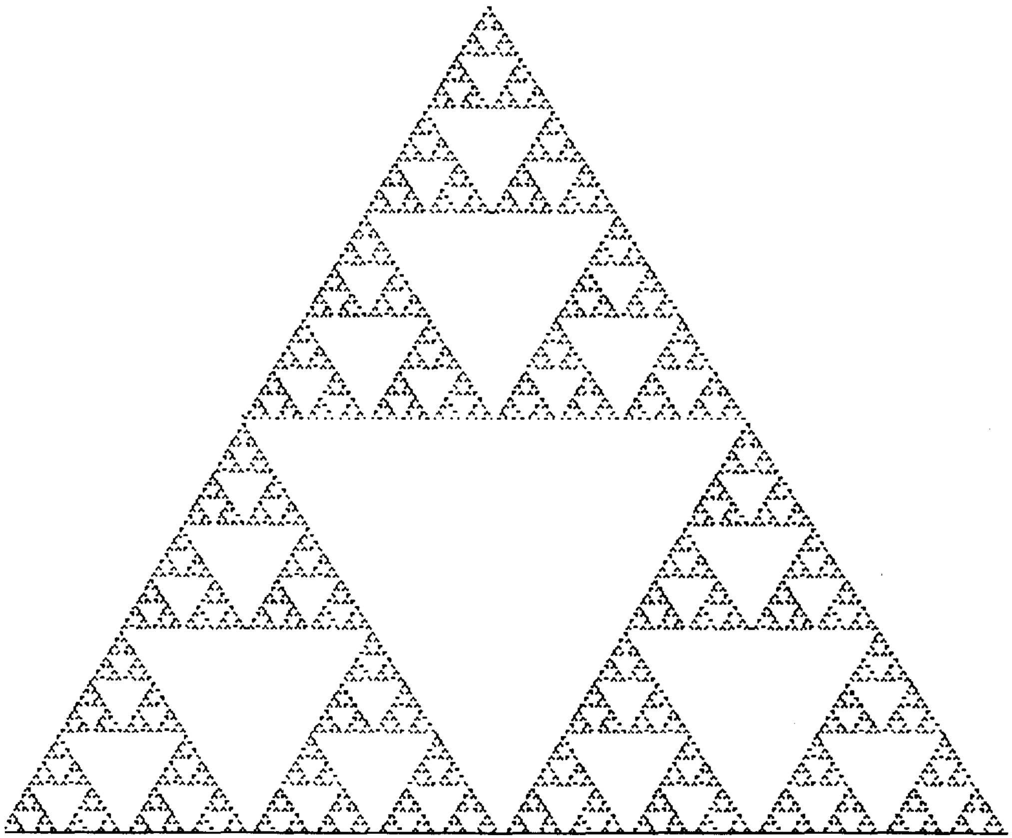


Figure 1: Sierpinski Triangle

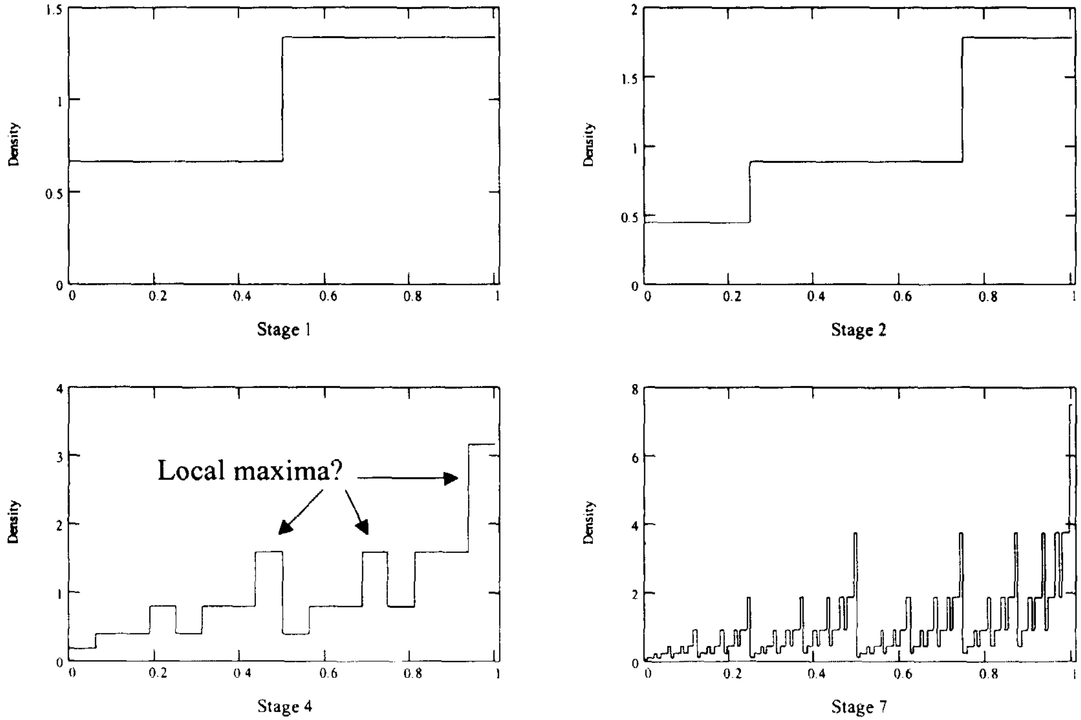


Figure 2: Stages of the Binomial Multifractal

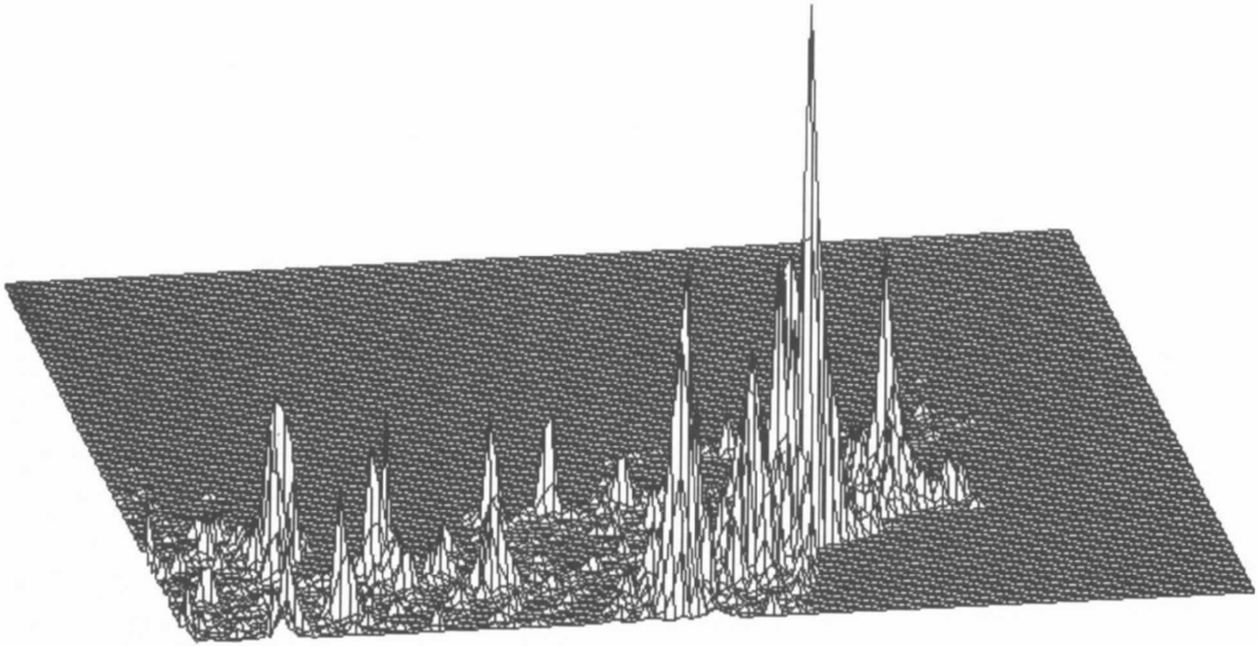


Figure 3: Northeastern USA Population Density

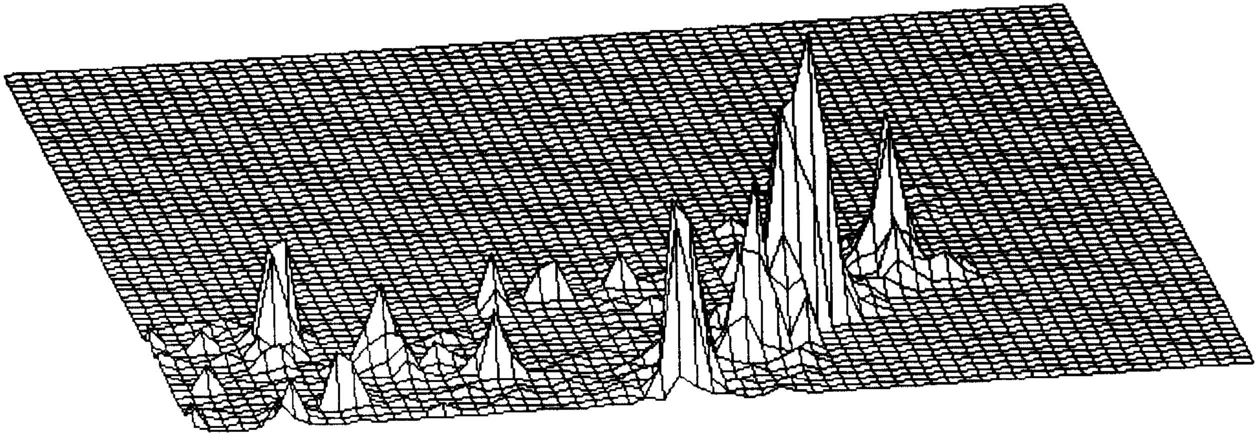


Figure 4: N.E. USA at 1/2 resolution

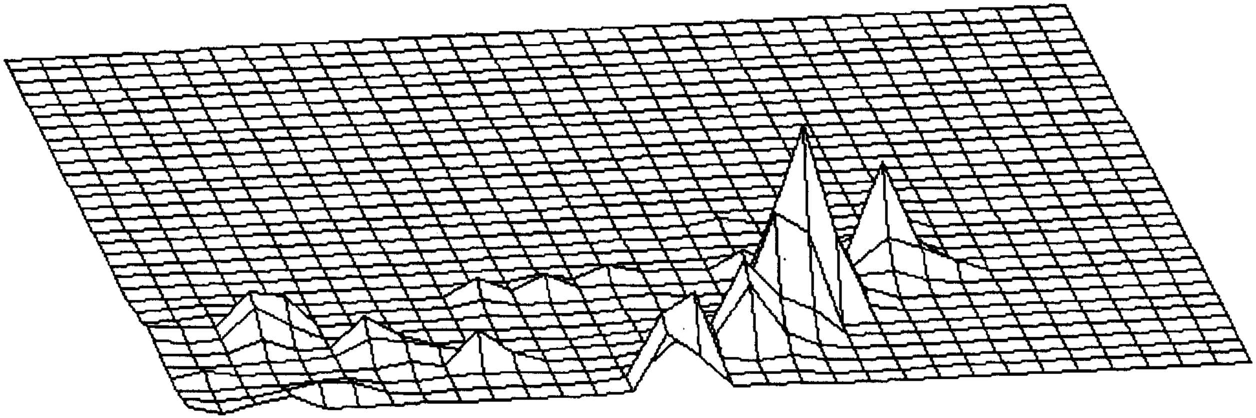


Figure 5: N.E. USA at 1/4 resolution

Figure 6: Sierpinski Multifractal

Figure 7: Random Sierpinski Multifractal

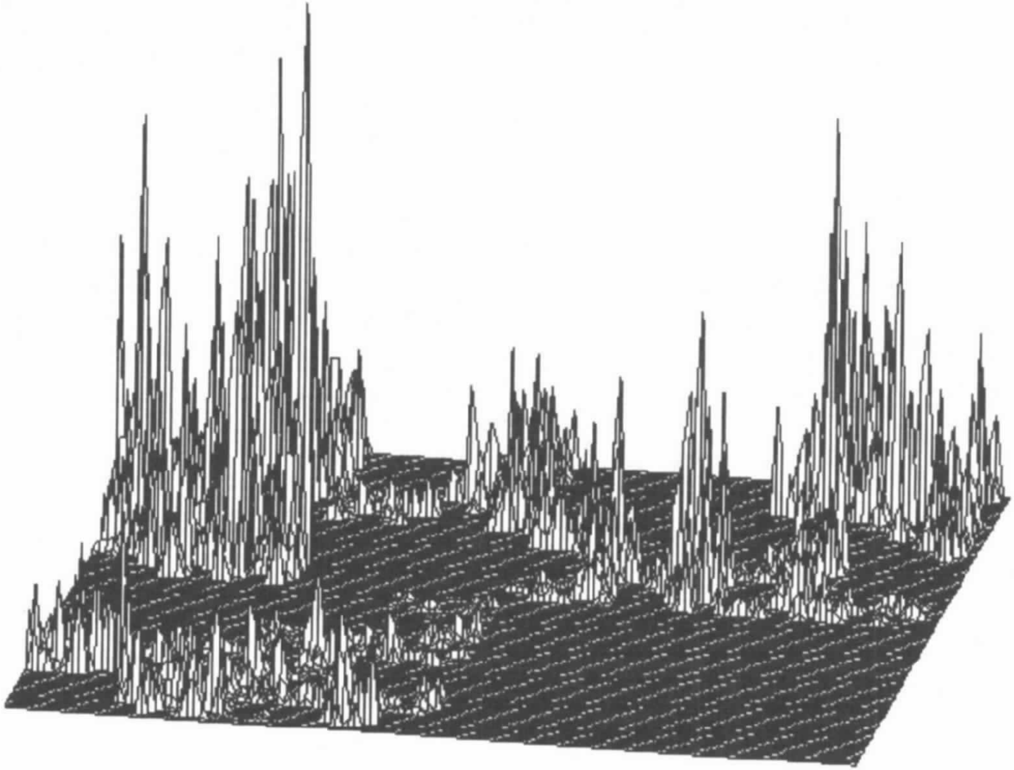


Figure 8: Random Sierpinski Multifractal (Perspective)

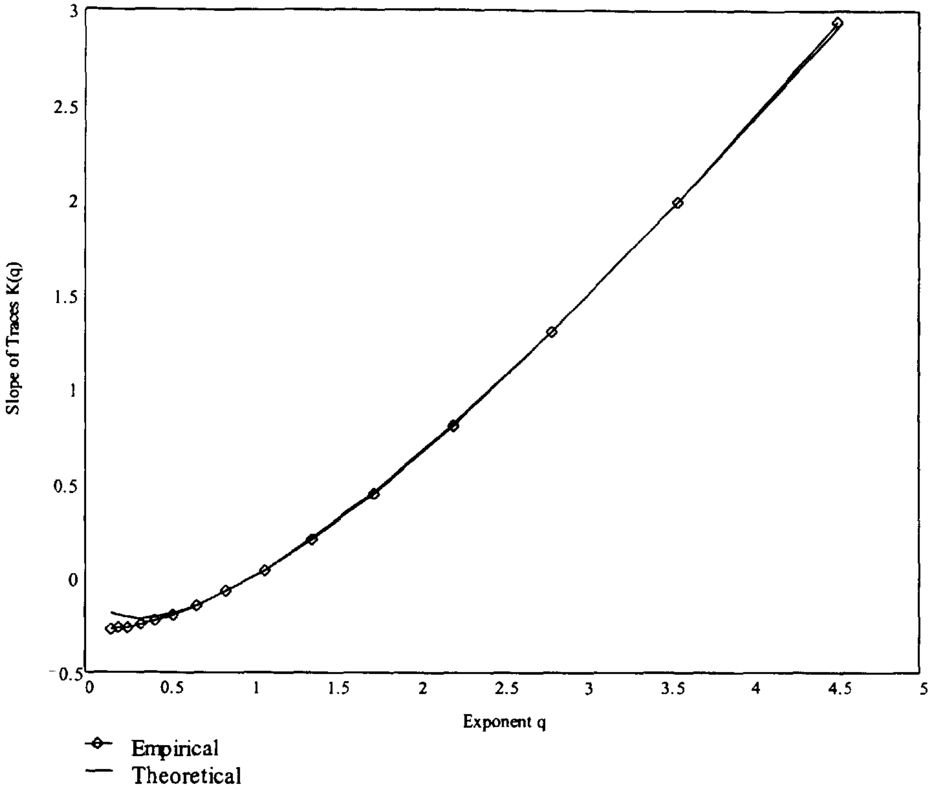


Figure 9: Sierpinski Multifractal K(q) Function

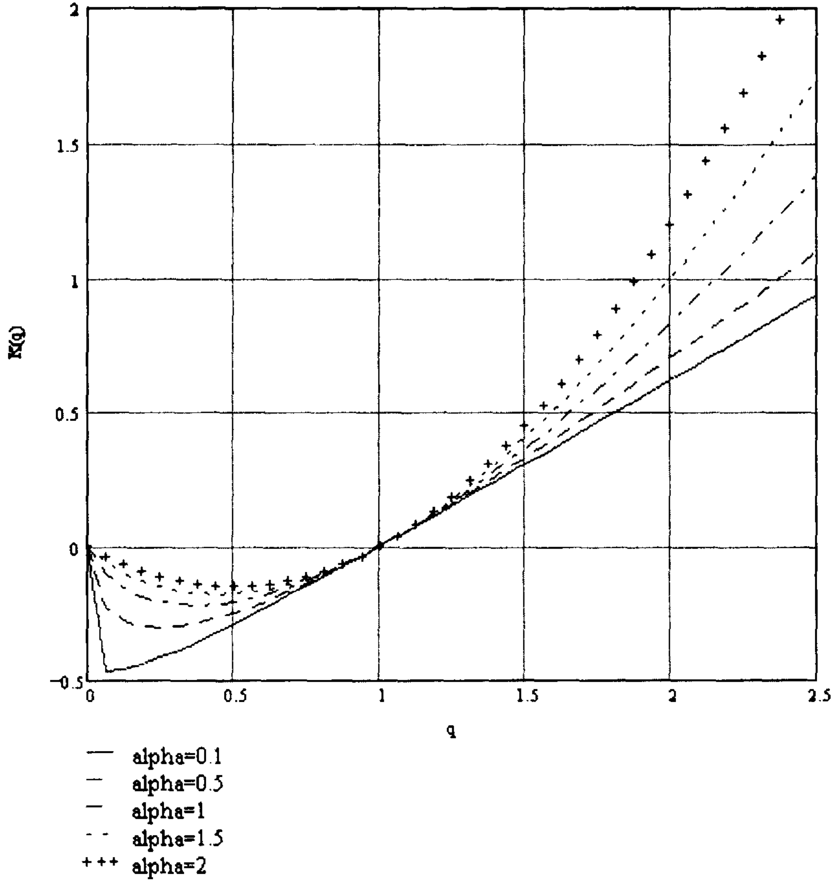


Figure 10: Effect of Alpha on K(q) Curve

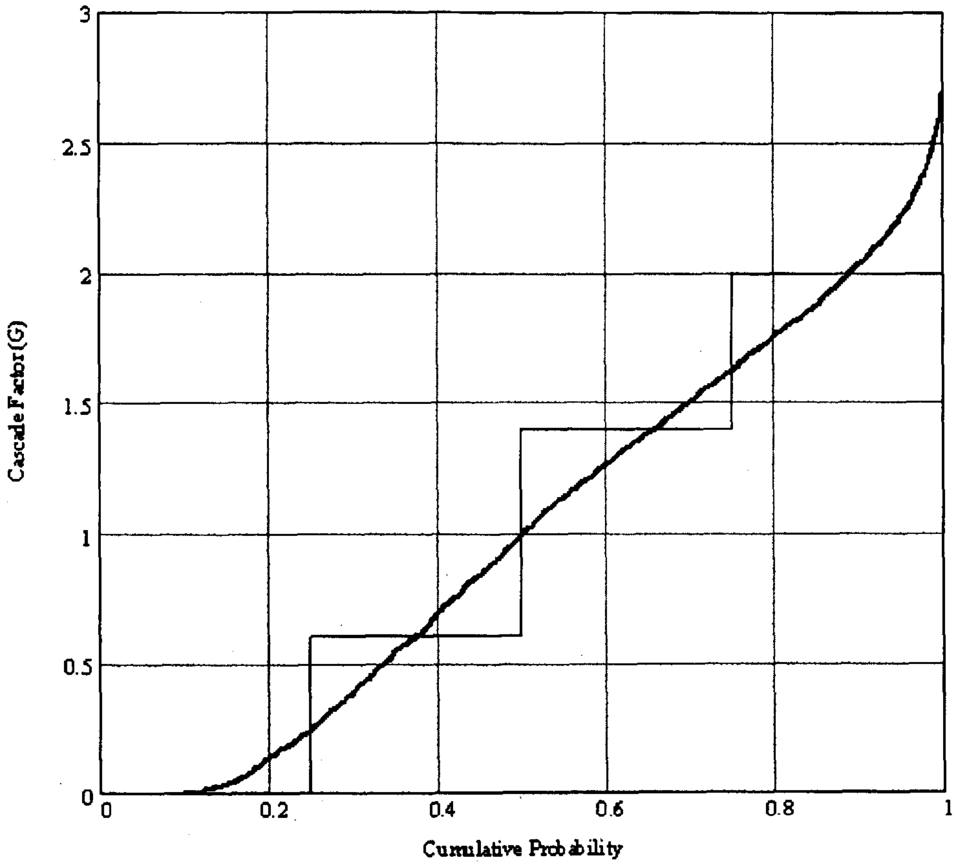


Figure 11: Universal Generator Distribution

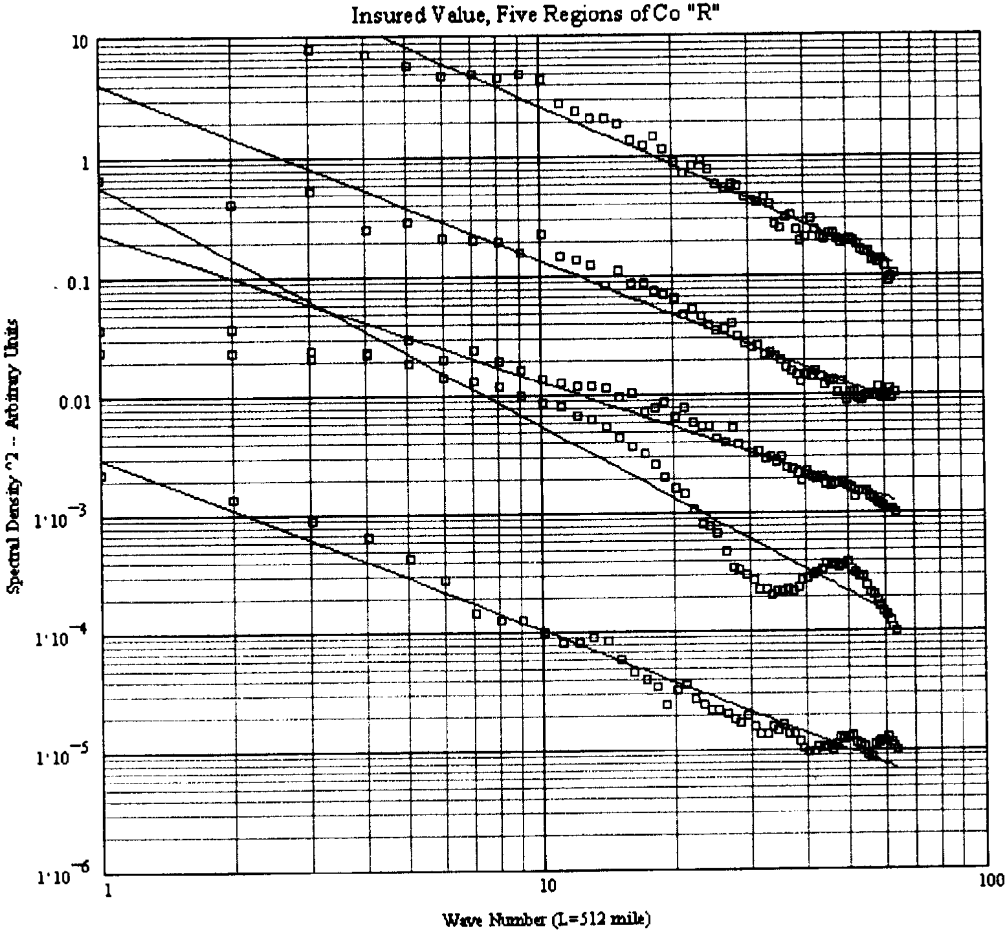


Figure 12: Comparative Portfolio Spectra

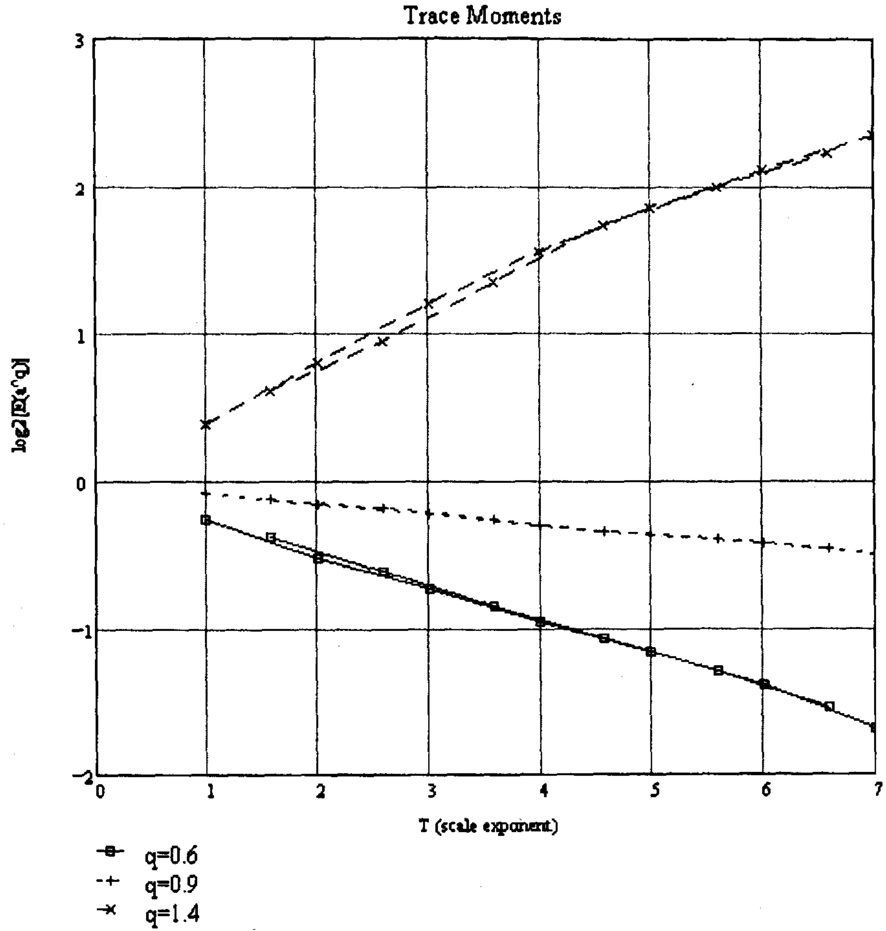


Figure 13: Selected Trace Moments of Population

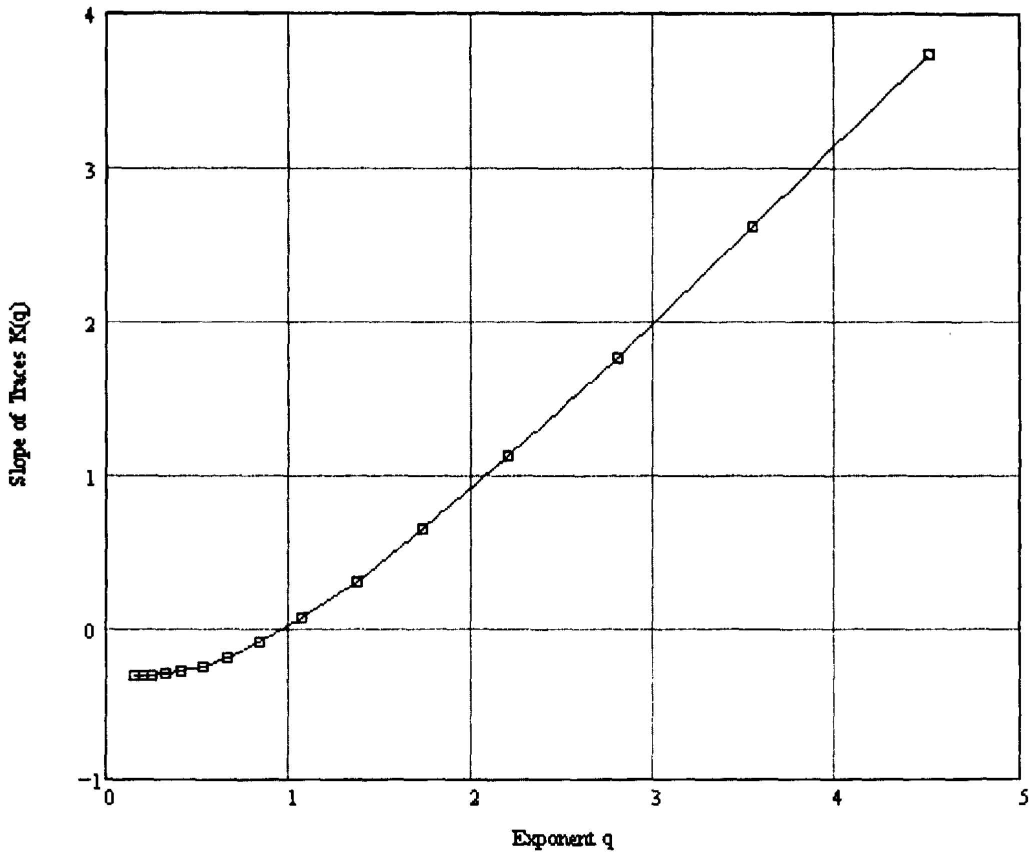


Figure 14: Empirical K(q) of Population

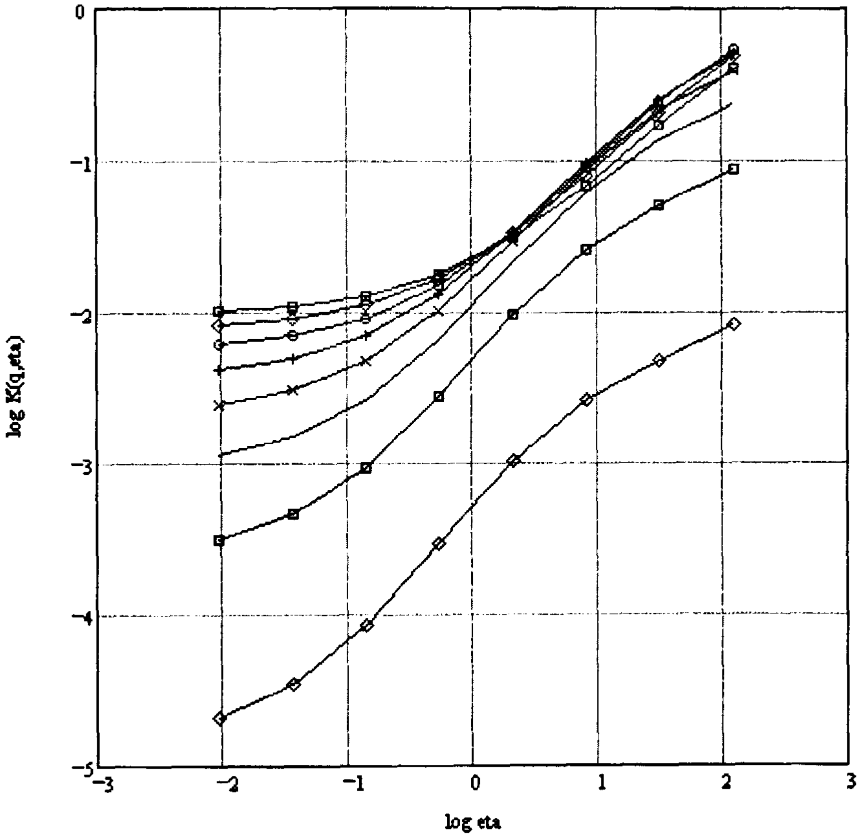


Figure 15: Double Trace Search for Alpha

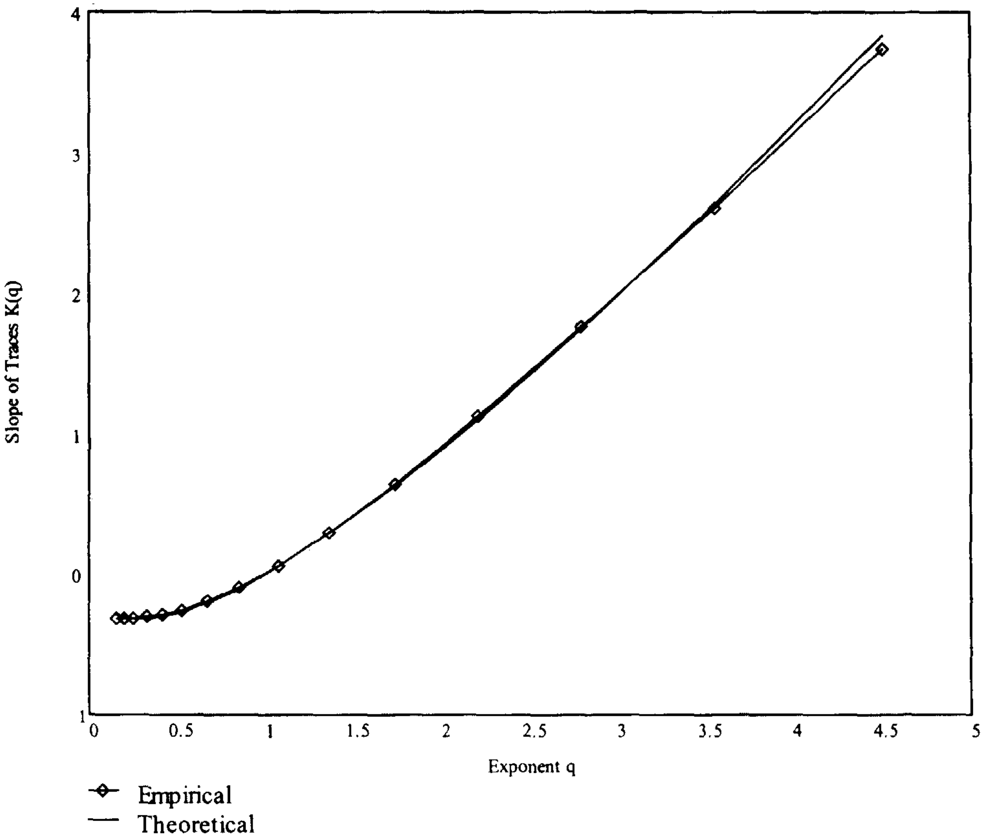


Figure 16: Empirical vs. Theoretical K(q) for Population

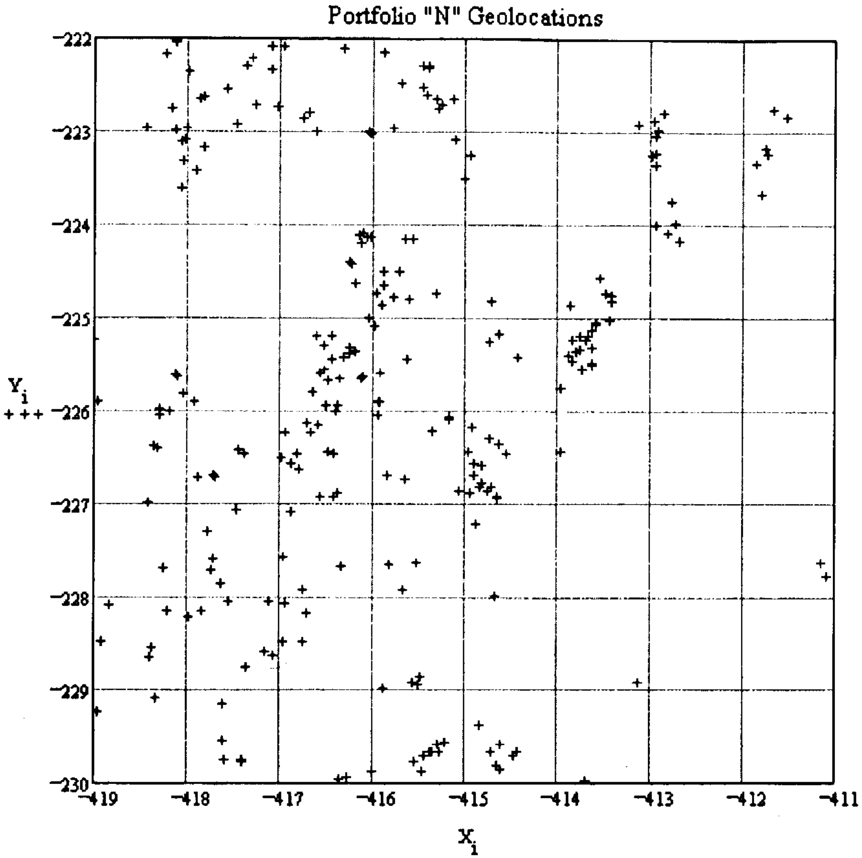


Figure 17: Actual Portfolio Map

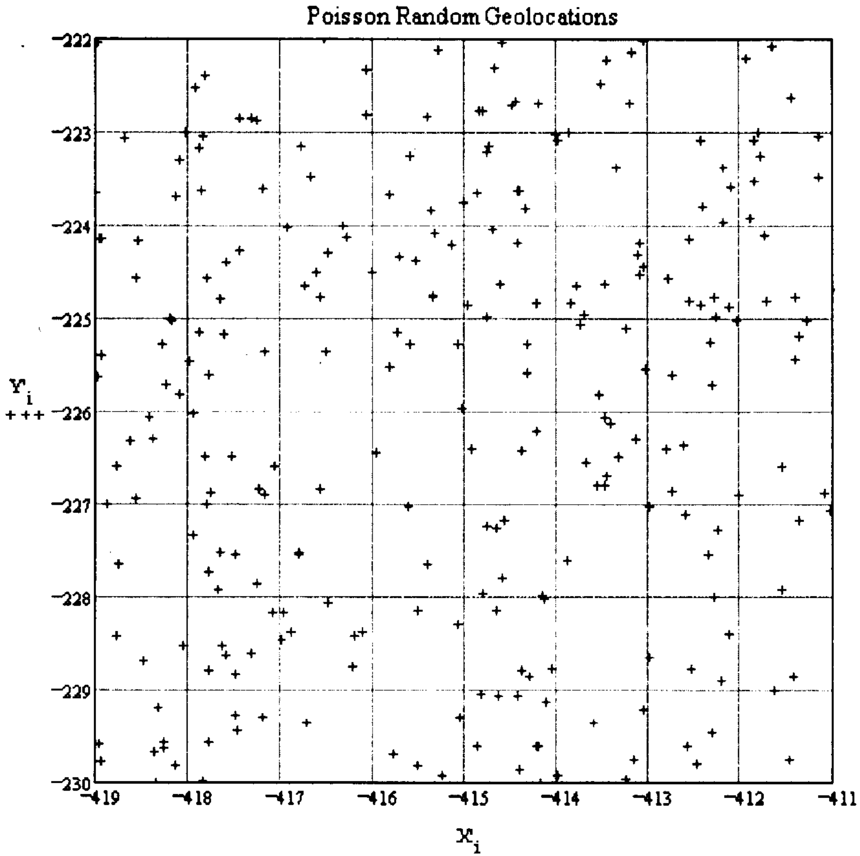


Figure 18: Poisson Simulation of Portfolio

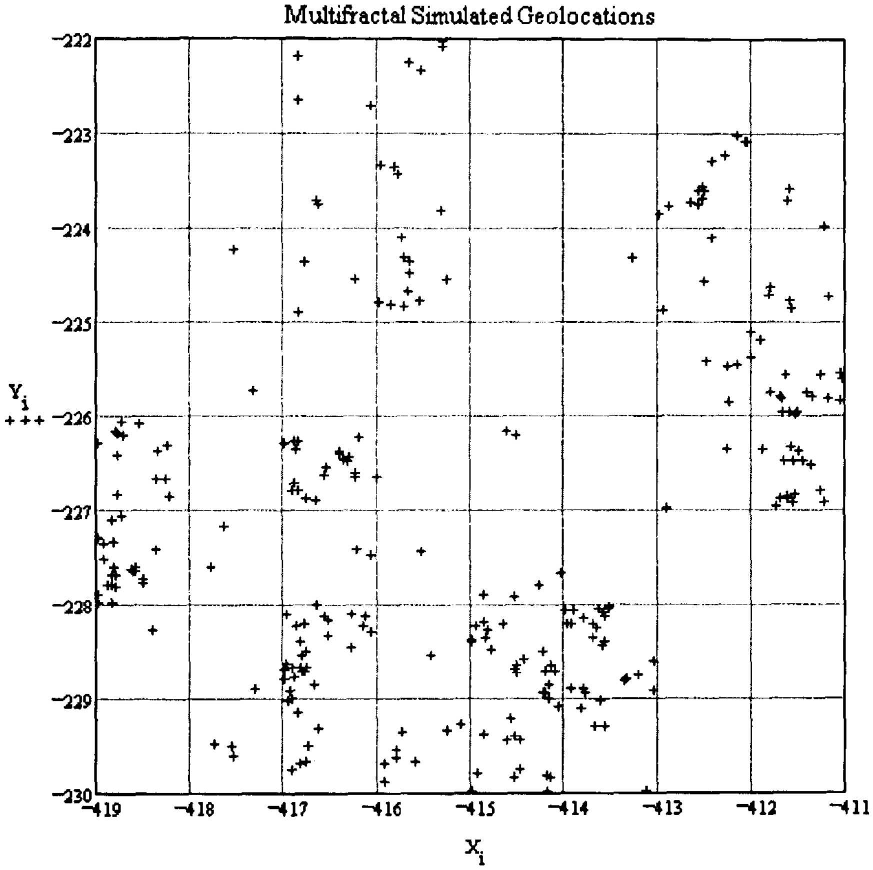


Figure 19: Multifractal Simulation of Portfolio

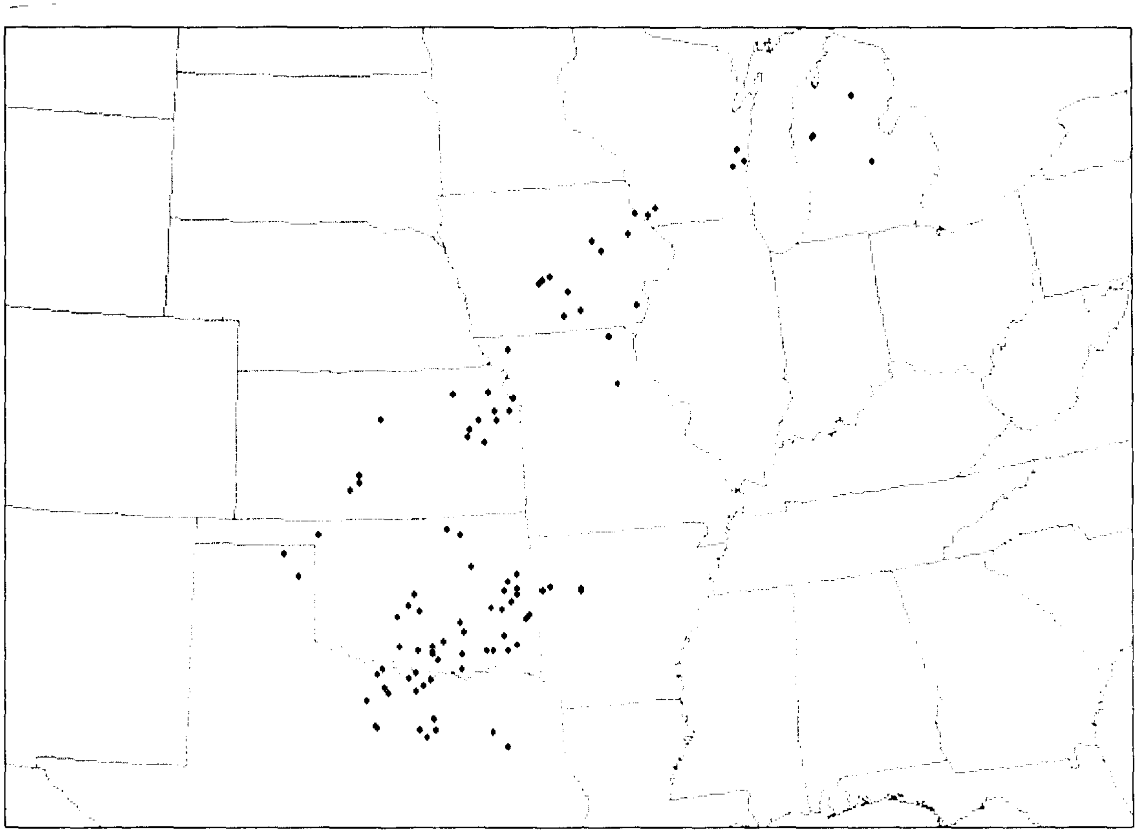


Figure 20: 3/30/98 Hail Reports As Given

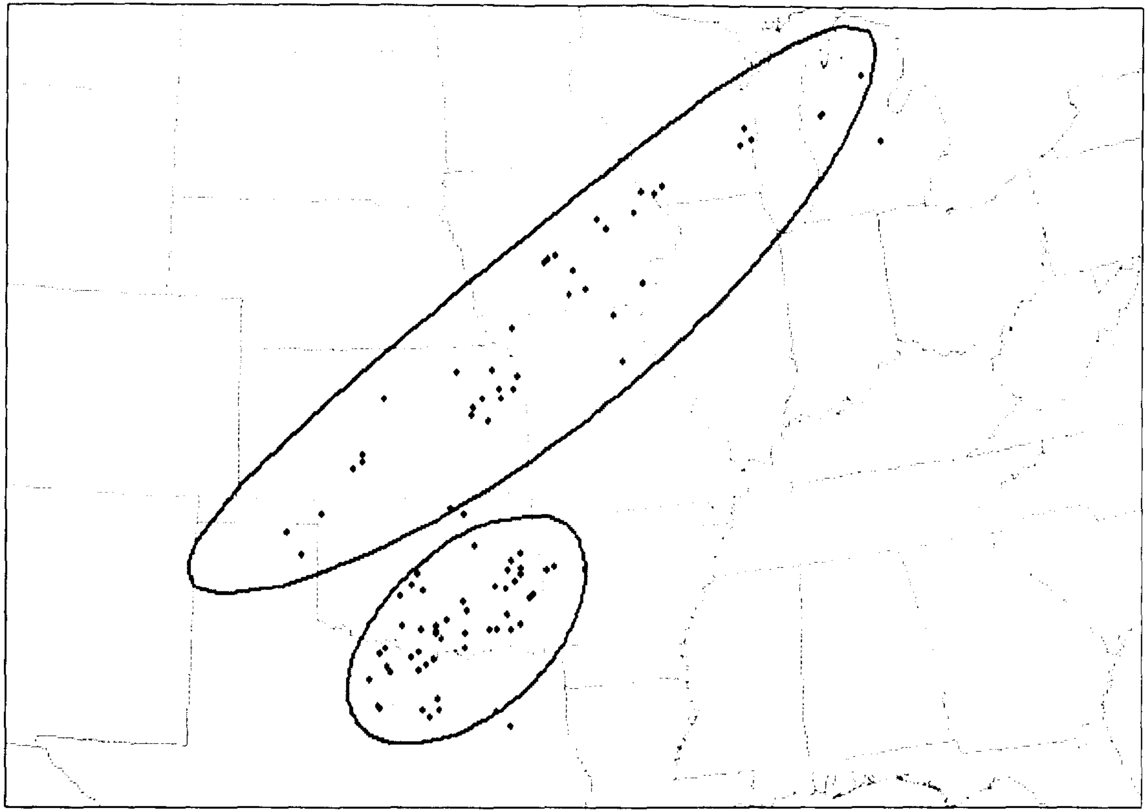


Figure 21: Hail Reports In Swaths

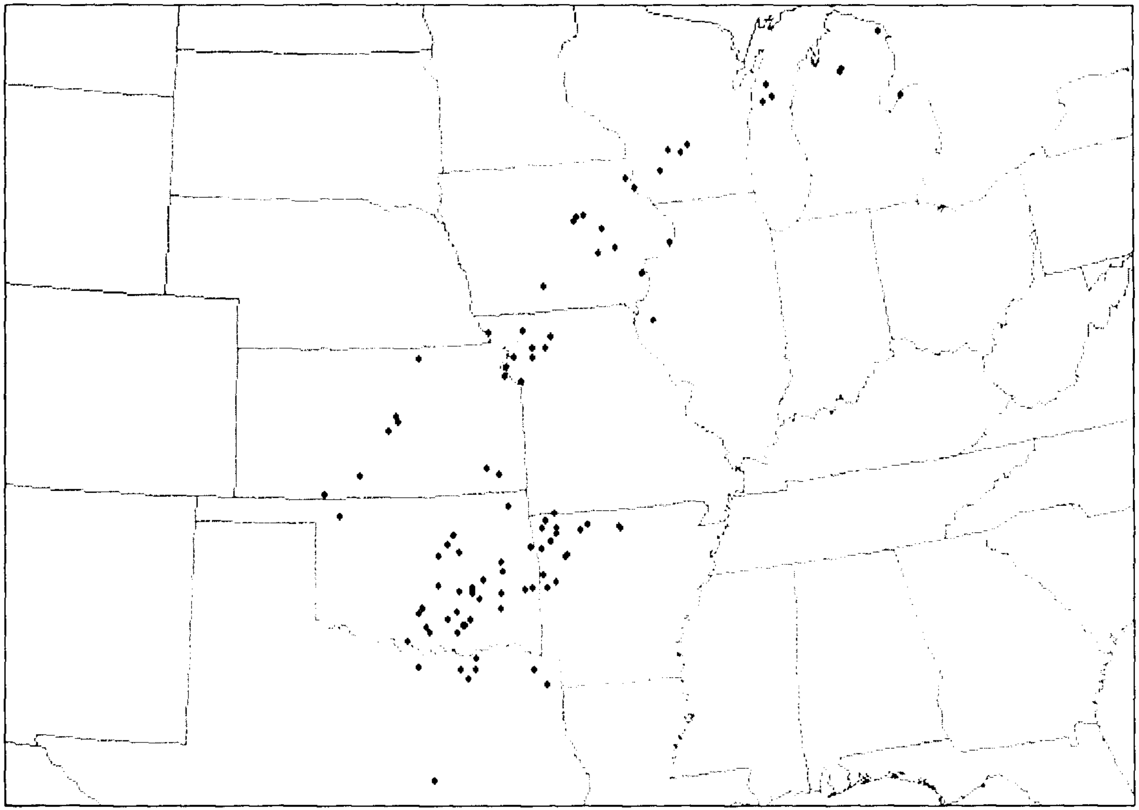


Figure 22: Hail Reports Shifted

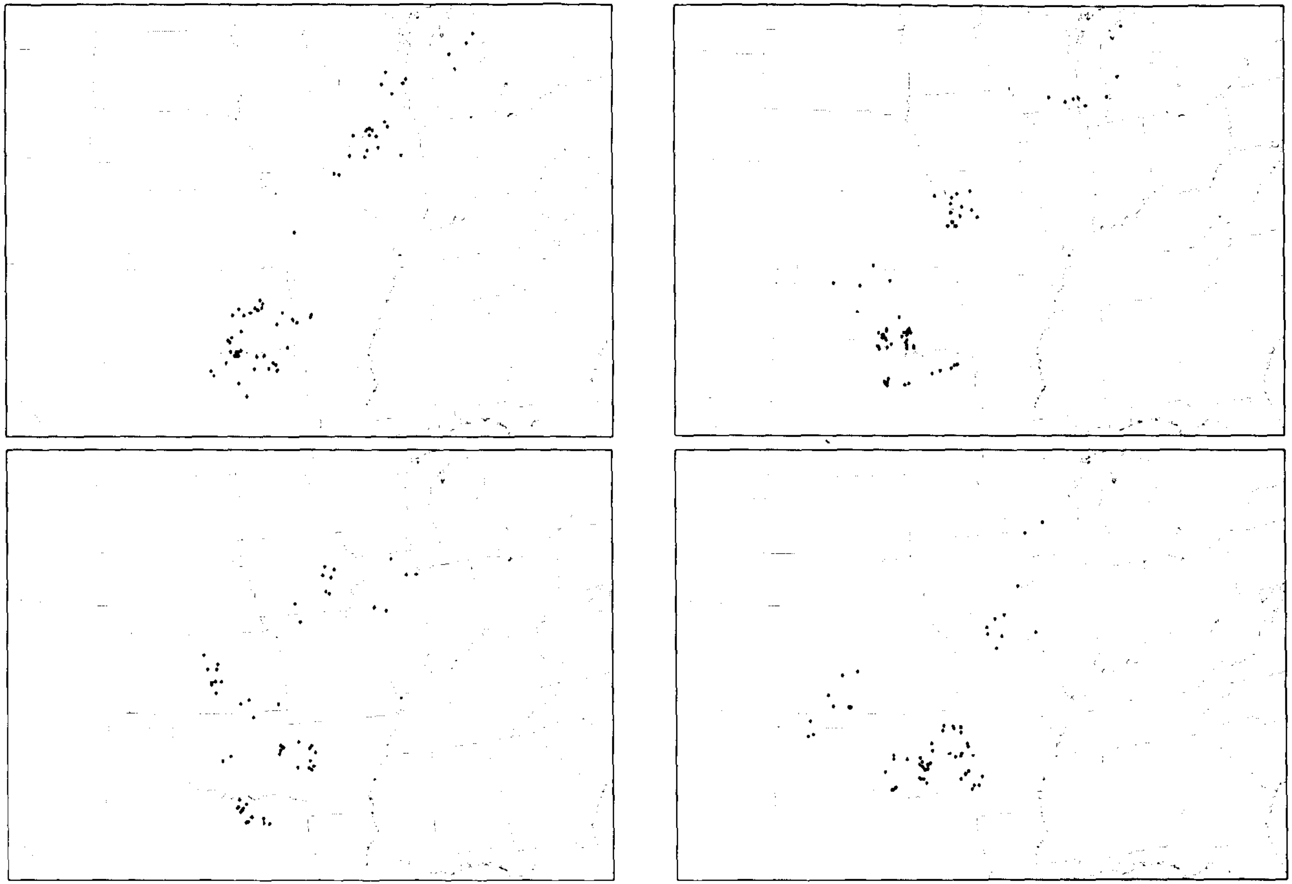


Figure 23: Multifractal Simulations of Hail Reports

Actuarial Applications of Multifractal Modeling Part II: Time Series Applications

Yakov Lantsman, Ph.D. and
John A. Major, ASA, MAAA

Actuarial Applications of Multifractal Modeling

Part II: Time Series Applications

by Yakov Lantsman, Ph.D. and John A. Major, ASA, MAAA

email:lantsman@netrisk.com, jmajor@guycarp.com

Abstract

Multifractals are mathematical generalizations of fractals, objects displaying "fractional dimension," "scale invariance," and "self-similarity." Many natural phenomena, including some of considerable interest to the casualty actuary (meteorological conditions, population distribution, financial time series), have been found to be well-represented by (random) multifractals. In part II of this paper, we show how to fit multifractal models in the context of one-dimensional time series. We also present original research on the multifractality of interest rate time series and the inadequacy of some state-of-the-art diffusion models in capturing that multifractality.

Introduction

In the accompanying part I paper, we introduced the ideas of fractal point sets and multifractal fields. We showed that those mathematical constructs are applicable to a wide range of natural phenomena, many of which are of considerable interest to the casualty actuary. We showed how to analyze sample data from multidimensional random fields, detect and measure multifractal behavior, fit a "universal" model, and use that model to simulate independent realizations from the underlying process. In particular, we discussed synthetic geocoding and the simulation of non-hurricane atmospheric perils.

The theory of self-similar random time series is more fully developed than the general multidimensional case. In this part II paper, we focus on time series analysis and financial applications. We present some additional theoretical machinery here and discuss applications to weather derivatives and financial modeling.

Time Series

Introduction to Multifractal Time Series Analysis; Structure Function

Financial and geophysical time series feature a large range of time scales and they are governed by strongly non-linear processes; this suggests the possible applicability of scaling (multifractal) models. We consider a random process $X(t)$ defined on the time segment $[0, T]$. The process $X(t)$ has variously represented exchange rates, interest rates, temperature and precipitation in our work.

As in the two-dimensional case, scale invariance is most readily tested by computing $P(k)$, the power spectrum of $X(t)$. In the case of a one-dimensional time series, standard techniques of spectral (Fourier) analysis are available in many off-the-shelf statistical and mathematical packages, including Microsoft EXCEL.

For a scaling process, one expects power law behavior:

$$P(k) \propto k^{-\beta} \quad (1)$$

over a large range of wave-numbers k (inverse of time). If $\beta < 1$, the process is stationary in the most accepted sense of the word [1], that is, $X(t)$ is statistically invariant by translation in t. If $1 < \beta < 3$, the process is non-stationary but has stationary increments and, in particular, the small-scale gradient (derivative or first difference) process will be stationary. Introducing the Hurst exponent H ($0 < H < 1$), a parameter describing the degree of stationarity of $X(t)$, we can express the exponent β as follows:

$$\beta = 2H + 1 \quad (2)$$

We can demonstrate a wide range of self-similar processes by changing the Hurst exponent: Brownian motion ($H = 0.5$, $\beta = 2$), an "anti-persistent" fractional Brownian motion ($0 < H < 0.5$, $1 < \beta < 2$), and "persistent" fractional Brownian motion ($0.5 < H < 1$, $2 < \beta < 3$). This is the class of additive models. The last has become popular for modeling financial time series.

Most of financial and geophysical time series demonstrate non-stationary behavior. This creates major complications if power spectrum analysis is the only available tool. It is well known [2] that knowledge of β alone is insufficient to distinguish radically different types of statistical behavior (the phenomena of "spectral ambiguity"). It is not so difficult to construct two processes with identical power spectra – one additive and sufficiently smooth, and the other one multiplicative with a high degree of intermittency. But such cases can be resolved with the help of multifractal analysis, which can be viewed as an extension in the time domain of scale-invariant spectral analysis.

An appealing statistical characteristic to use in exploring time series is the *structure function*. Structure function analysis of processes with stationary increments consists of studying the scaling behavior of non-overlapping fluctuations $\Delta X_\tau(t) = | X(t+\tau) - X(t) |$ for different time increments τ. One estimates the statistical moments of these fluctuations, which – assuming both scaling (1) and statistical translational invariance in time (i.e., the property of stationarity increments) – depend only on the time increment τ in a scaling way:

$$E(\Delta X_\tau(t)^q) \sim E(\Delta X_T{}^q)\left(\frac{\tau}{T}\right)^{\zeta(q)} \quad (3)$$

where $E(\Delta X_T{}^q)$ is a constant (T is the fixed largest time scale), $q > 0$ is the order of the moment, and $\zeta(q)$ is the scale invariant structure function. The expectation $E(\Delta X_\tau(t)^q)$ is assumed finite for q in an interval $[0, q_{max})$. The structure function $\zeta(q)$ is a focal concept in the one-dimensional theory of multifractals.

We examine some properties of $\zeta(q)$. By definition, we have $\zeta(0) = 0$. Davis A. et al. [1] show that $\zeta(q)$ will be concave: $d^2\zeta(q)/dq^2 < 0$. This is sufficient to define a "hierarchy of exponents" using $\zeta(q)$:

$$H(q) = \frac{\zeta(q)}{q} \quad (4)$$

It can also be shown that $H(q)$ is a non-increasing function. The second moment is linked to the exponent β as follows:

$$\beta = 1 + \zeta(2) = 2H(2) + 1 \quad (5)$$

Obtaining $\zeta(q)$ or, equivalently, $H(q)$ is the goal of structure function analysis. A process with a constant $H(q)$ function could be classified as "monofractal" or "monoaffine"; in the case of decreasing $H(q)$, multifractal or "multiaffine."

Additive processes can be shown to have linear $\zeta(q)$ or constant $H(q)$. For Brownian motion we have:

$$\zeta_{BM}(q) = \frac{q}{2} \quad (6)$$

For fractional Brownian motion (the fractional integration of order h of a Gaussian noise):

$$\zeta_{FrBM}(q) = q(h - \tfrac{1}{2}) \quad (7)$$

Note that Brownian motion corresponds to $h = 1$ (an ordinary integral of Gaussian white noise, which gives $H = ½$ in Fourier space).

In the case of the more exotic "Lévy flight" (additive processes with Lévy noise) the behavior of $\zeta(q)$ is still linear. In this case, there is a Lévy index α ($0 \le \alpha \le 2$), which characterizes the divergence of the moments of the Lévy noise. In general $\zeta(q)$ diverges for $q > \alpha$, but for finite samples we obtain the following $\zeta(q)$ function for a Lévy flight of index α:

$$\zeta_{LM}(q) = \frac{q}{\alpha} \quad (8)$$

for $q < \alpha$, and $\zeta(q) = 1$ for $q \ge \alpha$.

We thus see that observing non-linearity of an empirical $\zeta(q)$ function is a solid argument against the validity of an additive model. Below, we will show strong signs of curvature in the behavior of some empirical $\zeta(q)$ functions for financial time series.

The generic multifractal processes (non-linear, non-additive) could be modeled by multiplicative cascades. The central part of a multiplicative cascades is the generator (MCG, discussed in the part I paper) which should be represented by some infinitely divisible probability distribution. Using "canonical representation" (the Lévy-Khinchine representation) for infinitely divisible random variables, and arguments similar to those for the $K(q)$ function for the general D-dimensional case, we obtain the following "universal form" for the structure function of a non-stationary process:

$$\zeta(q) = qH - \frac{C1}{\alpha - 1}(q^{\alpha} - q) \quad (9)$$

where $H = \zeta(1)$ the same as (2), $C1$ is a parameter with the same role as in equation (31) of part I, and α is the Lévy index.

Analogous considerations could guide us to modify part I's equation (31) to express the *K(q)* function for a non-conservative field:

$$K(q)+qH=\begin{cases}\dfrac{C1}{\alpha-1}(q^{\alpha}-q) & \alpha\neq 1\\ C1q\log(q) & \alpha=1\end{cases} \tag{10}$$

where the H parameter is the degree of non-stationarity of the process. In other words, first bring the field to a state of stationarity (by fractional differentiation, i.e., power-law filtering in Fourier space or a small-scale gradient transformation) to eliminate the linear part qH, and then proceed with the analysis as for conservative fields.

To summarize, the basic steps are:

1. Examine the data for evidence of intermittency and self-similarity; this could be accomplished by studying the power spectrum.
2. Establish the status of multifractality (or monofractality) and qualitatively characterize the system under investigation; for this, we use the structure function.
3. Fit model parameters to the universal form of $\zeta(q)$.
4. Simulate, using multiplicative cascade techniques based on the universal form of the generator.
5. Apply, including, possibly, drawing inferences about the underlying process.

A Growing Crisis in Financial Time Series Modeling?

There is a growing awareness among researchers that the existing "classical" models cannot accommodate some essential properties underlying financial phenomena. The accumulation of a tremendous amount of highly reliable data from the financial markets around the world reveals distinctive characteristics of financial time series that had previously been overlooked because of lack of data. Some of the most important features are:

- scaling or self-similarity (at different time scales);
- long-term memory or persistency;
- volatility clustering;
- hyperbolic or "Paretian" tails.

To compensate for the consequences of these characteristics, the number of parameters in the "classical" models has been increasing over time. If this continues unchecked, it could make models unstable and decrease their predictive power.[1]

We distinguish two major classes of models in use by practitioners today: continuous time stochastic diffusion models ("diffusion models") and discrete time series models ("discrete models").

Diffusion models build on the well-understood theory of Brownian motion. The development of stochastic calculus (particular Itô integrals) and the theory of martingales created the essential mathematical apparatus for equilibrium theory. The assumption of arbitrage free pricing (rule of one price) has a very elegant mathematical interpretation as a change of stochastic measure and the transformation to a risk-neutral stochastic process.

Application of diffusion models is a crucial element in the valuation of a wide variety of financial instruments (derivatives, swaps, structured products, etc). Researchers have, however, long recognized major discrepancies between models based on Brownian motion and actual financial data, including long-term memory, volatility clustering and fat tails. To resolve these problems some extensions of diffusion models were offered. Often, this means introducing more stochastic factors, creating so-called multi-factor models.

Modem discrete models extend classical auto-regressive (AR) moving average (MA) models with recent advances in the parameterization of time-conditional density functions. These include ARCH, GARCH, PGARCH, etc. Discrete models have been partially successful in compensating for lack of long-term memory, volatility clustering and fat tails, but at the cost of an increasing number of parameters and structural equations. Using appropriate diagnostic techniques one can demonstrate that the statistical properties of discrete models (viz., self-similarity of moments, long-term memory, etc.) are essentially the same as for Brownian motion.

There is a third class of models, in little use by practitioners, but familiar to academics. This group constitutes the so-called additive models, including fractional Brownian motion, Lévy flight and truncated Lévy flight models. These models can replicate *mono*-fractal structure of underlying processes – their corresponding structure functions $\zeta(q)$ (7), (8) are linear – but they cannot produce *multi*fractal (nonlinear) behavior.

Case Study: Foreign Exchange

Here, we present an example of the application of multifractal analysis to exchange rate modeling, substantially following Schmidt, F. [3]. Figure 1 represents the exchange rate time series (US$/GDM spot rate 1975 - 1990 weekly observations) and Figure 2 the corresponding logarithmic changes in exchange rate.

Figure 3 represents a power spectrum analysis (in log-log space) of the FX time series. Visual inspection, and the close fit of the regression line, supports the hypothesis of scale-invariant behavior. The power spectrum obeys a power law (Equation 1). The slope

[1] A similar "Ptolemaic crisis" afflicted meteorological precipitation modeling in the 1980s. See, e.g., the Water Resources Research special issue on Mesoscale Precipitation Fields, August 1985.

of the straight line is the parameter β; here equal to 1.592. This value suggests the underlying process may be non-stationary but with stationary increments.

An important application of multifractal analysis is to characterize *all* order moments for the validation of a scaling model. The appropriate tool to do this for the particular case of a time series is structure function analysis.

To apply the structure function method, we rewrite the equation (3) in logarithmic form:

$$\log[E(\Delta X_\tau(t)^q)] = \log[E(\Delta X_T{}^q)] + \zeta(q)\{\log(\tau) - \log(T)\} \qquad (11)$$

The expectation $E(\Delta X_\tau(t)^q)$ is estimated by the so-called *partition function*

$$E(\Delta X_\tau(t)^q) \cong \frac{1}{N}\sum_\tau |\Delta X_\tau(t)|^q \qquad (12)$$

(see Fisher, A. et al. [4]). We then plot $\log[E(\Delta X_\tau(t)^q)]$ against $\log(\tau)$ for various values of q and various values of τ. Linearity of these plots for given values of q indicates self-similarity. Linearity could be checked by visual inspection or by some more sophisticated techniques (e.g., significance test for higher-order regression terms). The slope of the line, estimated by least squares regression, gives an estimate of the scaling function $\zeta(q)$ for that particular q.

The structure function, mapping q to its slope, is depicted in Figure 4. Here, we also draw an envelope of two straight lines corresponding to Brownian motion (slope 0.5) and fractional Brownian motion (slope 0.6), respectively. The non-linear shape of the empirical curve is the signature of multifractality.

Having established the existence of multifractality in the data, we can move to the next step – fitting parameters. In the case of one dimensional (time series) field, we use equation (9) to find universal parameters. For FX data, the universal parameters are: H = 0.532, α = 1.985, $C1$ = 0.035.

Case Study: Interest Rates

In this section, we present an original analysis of US interest rates. We use weekly observations of 3-month Treasury Bill Yield Rates (1/5/1962 - 3/31/1995). Figure 5 represents the interest rate time series and Figure 6 the corresponding logarithmic changes in interest rate from one period to the next.

We start with the power spectrum in Figure 7. Visual inspection and regression confirm the hypothesis of scaling behavior with corresponding β_{eir} = 1.893. The value of β_{eir} indicates that this interest rate series *might* be modeled by a non-stationary process with stationary increments.

Figure 8 represents the $\zeta(q)$ curve for interest rates with the same Brownian motion and fractional Brownian motion lines that we used for the FX analysis overlaid on the graph for reference. Again, the signature of multifractality is clearly present in the data. We obtain the following universal values: H = 0.612, α = 1.492, $C1$ = 0.095. These values could be used to simulate interest rates by applying a multiplicative cascade technique.

Andersen-Lund is Not Multifractal

We present an original analysis of the three-factor Anderson-Lund model of interest rates and show that even this model, with its highly complex structural equations and difficult fitting techniques, cannot replicate key features of empirical interest rate data.

The general form of the diffusion model Vetzal, K. [5] is

$$dX = \mu(X,t)\cdot dt + \sigma(X,t)\cdot dW \tag{13}$$

where X is the (possibly vector) random variable evolving over time, μ is a (possibly vector) function describing the instantaneous rate of change of X at a point in time, and σ is a (possibly matrix) function describing the instantaneous impact of changes in the (possibly vector) Gaussian random walk W, i.e. dW is a (are independent) Gaussian random variable(s). In the multidimensional case, the dimension of X does not necessarily equal the dimension of W. For interest rate models, one of the elements of X will represent the short rate of interest.

The primary purpose of these models has been to develop arbitrage-free prices of illiquid bonds, interest rate derivatives, etc., so the primary consideration has been fidelity in reproducing available market data, in particular, yield curves. However, obtaining realistic depictions of the objective behavior of the short rate (the historical evolution of the short rate over time) has been an important secondary consideration. It is this tension between the desire for analytical tractability on the one hand and realism on the other that has driven the development of ever more sophisticated models.

The simplest such model is Merton [6]:

$$dr_t = \theta\cdot dt + \sigma\cdot dW \tag{14}$$

where r_t is the interest rate at time t, θ is the average growth rate of the process, and σ is a volatility scale parameter.

Perhaps the most sophisticated of the analytically tractable models is the Cox-Ingersoll-Ross (CIR) model [7]:

$$dr_t = \kappa\cdot(\theta - r_t)dt + \sigma\cdot r_t^{1/2}\cdot dW \tag{15}$$

where κ is the mean reversion constant and θ is the global mean of the process. CIR adds realism to the Merton model by introducing mean reversion and volatility that is functionally dependent on the level of the rate.

Visual analysis of the interest rate time series graphs (as well as statistical diagnostics) reveals several distinctive features to US interest rates that cannot be accommodated by the CIR model.

1. Local trends in interest rate movements, indicating a changing mean to which the process reverts.
2. Heteroscedasticity that is not simply a function of the level of the rates.
3. Volatility clustering.

To address these limitations of CIR and previous models, Andersen and Lund [8] introduced the following (analytically intractable) three-factor model:

$$dr_t = \kappa_1 \cdot (\mu_t - r_t)dt + \sigma_t \cdot r_t^{\gamma} \cdot dW_{1,t} \tag{16}$$

$$d\log\sigma_t^2 = \kappa_2 \cdot (\alpha - \log\sigma_t^2)dt + \xi_1 \cdot dW_{2,t} \tag{17}$$

$$d\mu_t = \kappa_3 \cdot (\theta - \mu_t)dt + \xi_2 \cdot \mu_t^{1/2} \cdot dW_{3,t} \tag{18}$$

where r_t is the interest rate at time t, σ_t is the (unobserved) volatility, μ_t is the (unobserved) local mean of the process, k_1, k_2, k_3 are mean reversion constants, θ is the global mean of the process, α is the global mean of the log-volatility process, γ, ξ_1, ξ_2 are parameters, and $W_{1,t}$, $W_{2,t}$, $W_{3,t}$ are independent Gaussian random variables. Equation (16) can be seen as a generalization of the CIR model with unknown parameter γ instead of ½, and the fixed mean and volatility terms being replaced by endogenous variables evolving through their own diffusions (Equations 17 and 18).

The Andersen-Lund model represents the most realistic diffusion model we are currently aware of. The price of its realism is the need for substantial computing power. To fit the parameters, Andersen and Lund use the so-called Efficient Method of Moments procedure Gallant and Tauchen [9], which is an iterative method involving many simulations of the diffusion process. Calculation of yield curves similarly requires many simulation cycles, as there is no (known) closed-form solution.

Just how realistic is it? Figure 9 shows a 5,000-quarter simulation of interest rates using the A-L model with their recommended parameters. Figure 10 is the corresponding logarithmic changes in interest rate. A visual comparison of these graphs with the corresponding empirical interest rate graphs, and a cursory statistical examination of same, seems to validate the A-L approach to modeling interest rate time series. We will demonstrate that a deeper analysis of the scaling properties of *all* moments (not just the first and second) reveals fundamental differences between the A-L simulation and the empirical data. The simulated A-L data does not exhibit the multifractality that real interest rates possess.

Figure 11 shows the power spectrum function of the simulated time series out of the A-L model. Here, the parameter $\beta_{sir} = 1.772$, which is fairly close to the value obtained for the empirical data. The power spectrum, however, represents second moment statistics only. Its slope is not sufficient to validate a particular scaling model: it gives only partial information about the statistics of the process. One would need full knowledge of the probability distribution of the process or, equivalently, all of its statistical moments (not just second order) for a full validation.

Figure 12 represents the $\zeta(q)$ curve for the A-L simulated interest rate series, with the usual Brownian motion and fractional Brownian motion lines drawn for reference. Visual inspection and statistical testing indicate that the structure function of the data simulated by the A-L model and that of Brownian motion are nearly identical; the stochastic process underlying the A-L model appears to be *mono*fractal.[2]

The fundamental difference in scaling behavior revealed by the structure function comparison could lead to qualitatively different time series behavior. The universal parameters fit to the empirical process in the previous section indicate that the underlying mechanism should have a multiplicative cascade structure with (approximate) Lévy generator, rather than an additive process of information accumulation (Brownian motion type). Paraphrasing Müller et al. [10], the large scale volatility predicts small scale volatility much better than the other way around. This behavior can be compared to the energy flux in hydrodynamic turbulence, which cascades from large scales to smaller ones, not vice-versa.

Conclusions

In the companion part I paper, we introduced the ideas of fractal point sets and multifractal fields. We showed that while those mathematical constructs are rather bizarre from a traditional point of view (e.g., theory of smooth, differentiable functions), they nonetheless have applicability to a wide range of natural phenomena, many of which are of considerable interest to the casualty actuary. We showed how to analyze sample data from multidimensional random fields, detect scaling through the use of the power spectrum, detect and measure multifractal behavior by the trace moments and double trace moments techniques, fit a "universal" model to the trace moments function K(q), and use that model to simulate independent realizations from the underlying process by a multiplicative cascade. In particular, we discussed synthetic geocoding and the simulation of hail and tornadoes.

In this part II paper, we showed how to analyze time series through the structure function, and showed particular examples of foreign exchange and interest rate time series. We discussed the variety of time series models in use by practitioners and theoreticians and showed how even state-of-the-art diffusion models are not able to adequately reflect the multifractal behavior of real financial time series.

The field of stochastic modeling is constantly growing and evolving, so the term "Copernican revolution" might be too strong to describe the advent of multiplicative cascade modeling. Nonetheless, multifractals have clearly taken hold in the realm of geophysical and meteorological modeling, and it seems clear that they will eventually find their place in the world of financial models, as well. However, there are still numerous open questions, such as how to implement arbitrage-free pricing, that need to be answered before multifractal models can replace diffusion models as explanations of market pricing mechanisms.

[2] Theoretical arguments suggest monofractality for any additive models, Schmidt [3].

References

1. A. Davis, A. Marshak, W. Wiscombe, and R. Cahalan, "Multifractal characterizations of nonstationarity and intermittency in geophysical fields: Observed, retrieved, or simulated," *Journal of geophysical research,* Vol. 99, N. D4, pp.8055-8072, April 20, 1994.
2. D. Schertzer, and S. Lovejoy, "Physical modeling and analysis of rain clouds by anisotropic scaling multiplicative processes," *Journal of geophysical research,* Vol. 92, pp. 9693-9714, 1987.
3. F. Schmitt, D. Schertzer, and S. Lovejoy, "Multifractal analysis of foreign exchange data," submitted to *Applied Stochastic Models and Data Analysis (ASMDA).*
4. A. Fisher, L. Calvet, and B. Mandelbrot, "Multifractality of deutschmark / US dollar exchange rates," *Cowles Foundation Discussion Paper* # 1165, 1997.
5. K. Vetzal, "A survey of stochastic continuous time models of the term structure of interest rates," *Insurance: Mathematics and Economics* #14, pp. 139-161, 1994.
6. R. C. Merton, "Theory of rational option pricing," *Bell Journal of Economics and Management Science,* Vol. 4, pp. 141-183, 1973.
7. J. Cox, J. Ingersoll, and S. Ross, "A theory of the term structure of interest rates," *Econometrica* # 53, pp. 385-407, 1985.
8. T. Andersen, and J. Lund, Stochastic "Volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve," *Working Paper* #214, 1996.
9. A. Gallant, and G. Tauchen, "Estimation of continuous time models for stock returns and interest rates," *Manuscript,* Duke University, 1995.
10. U. Müller, M. Dacorogna, R. Dave, R. Olsen, O. Pictet, and J. von Weizsacker, "Volatilities of different time resolutions – Analyzing the dynamics of market components," *J. of Empirical Finance* # 4, pp. 213-239, 1997.

Figures for Part II

Yakov Lantsman
John A. Major

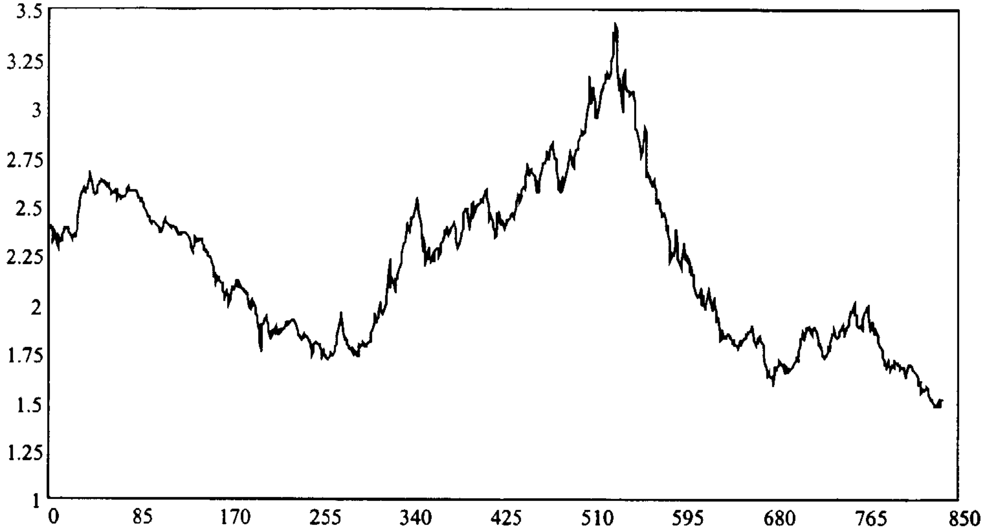


Figure 1: US$/GDM Exchange rate time series

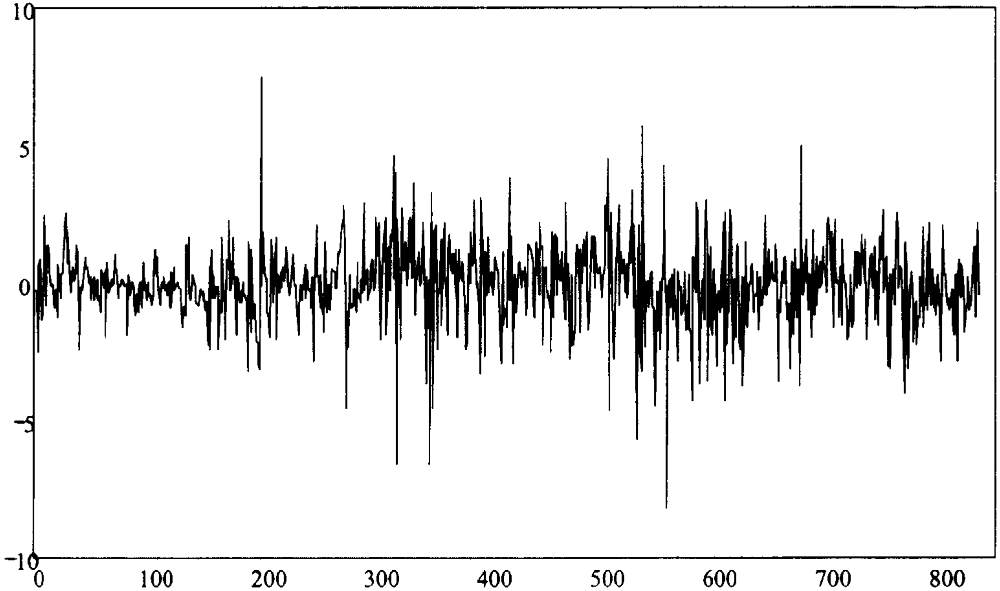


Figure 2: Logarithmic changes of Figure 1

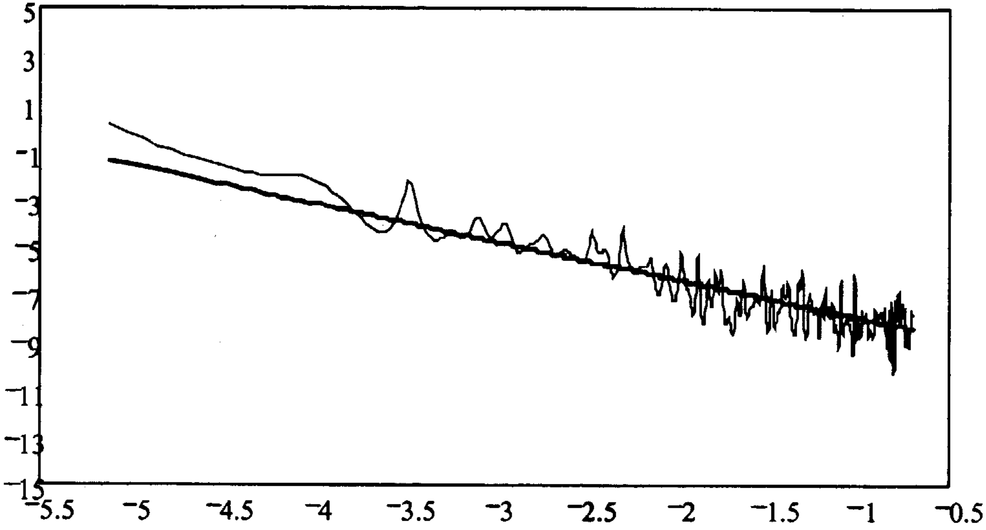


Figure 3: Power spectrum of FX data (log-log)

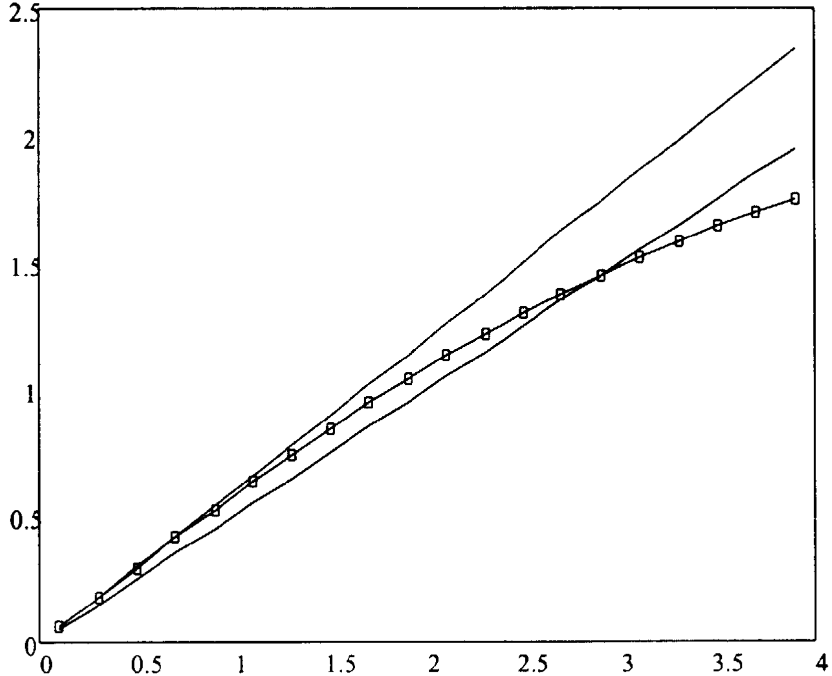


Figure 4: Structure Function Curve for FX

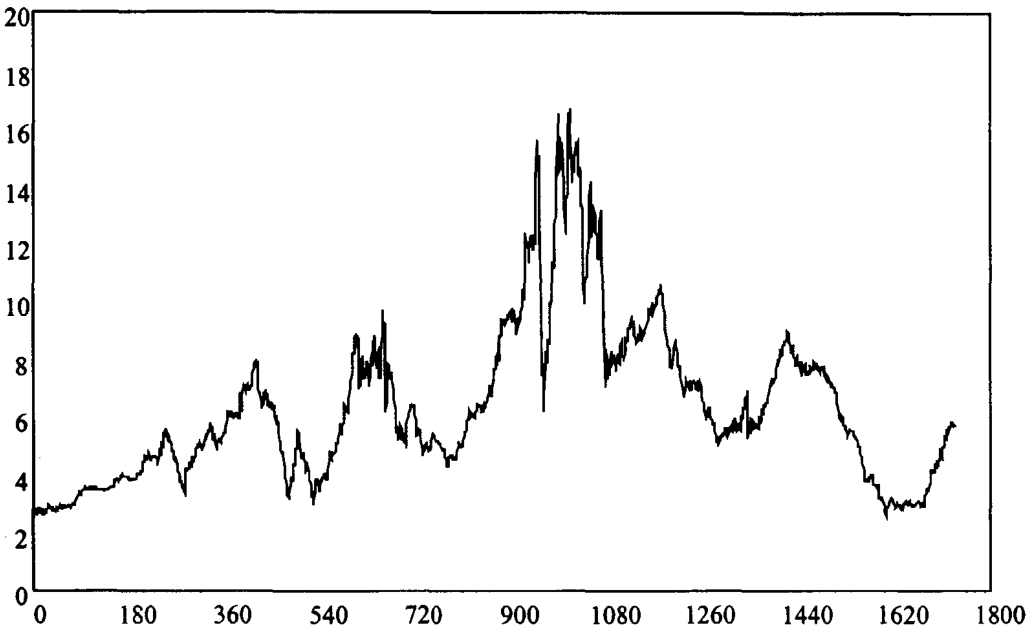


Figure 5: 3-mo T-Bill rates

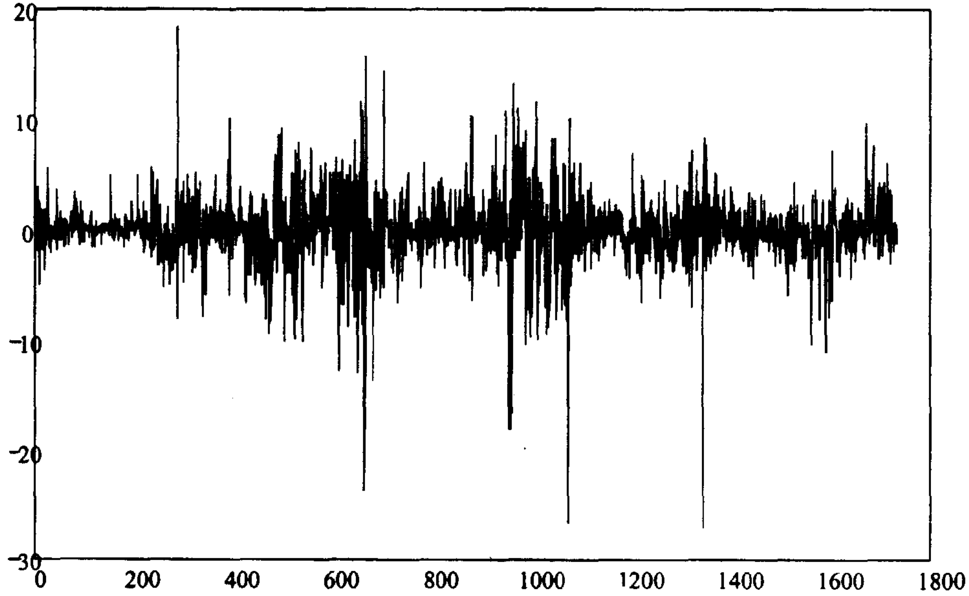


Figure 6: Logarithmic changes of Figure 5

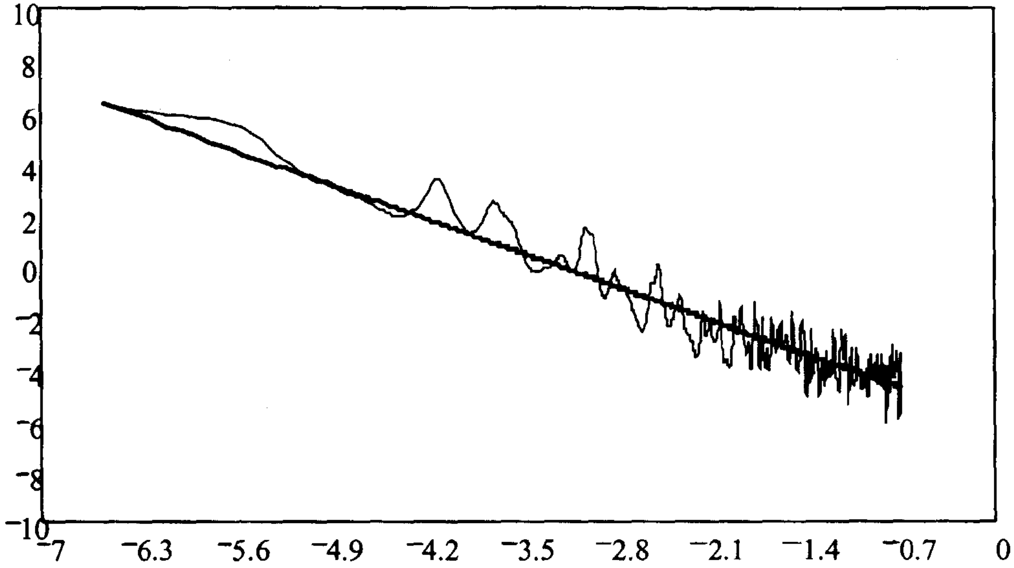


Figure 7: Power spectrum of interest rate

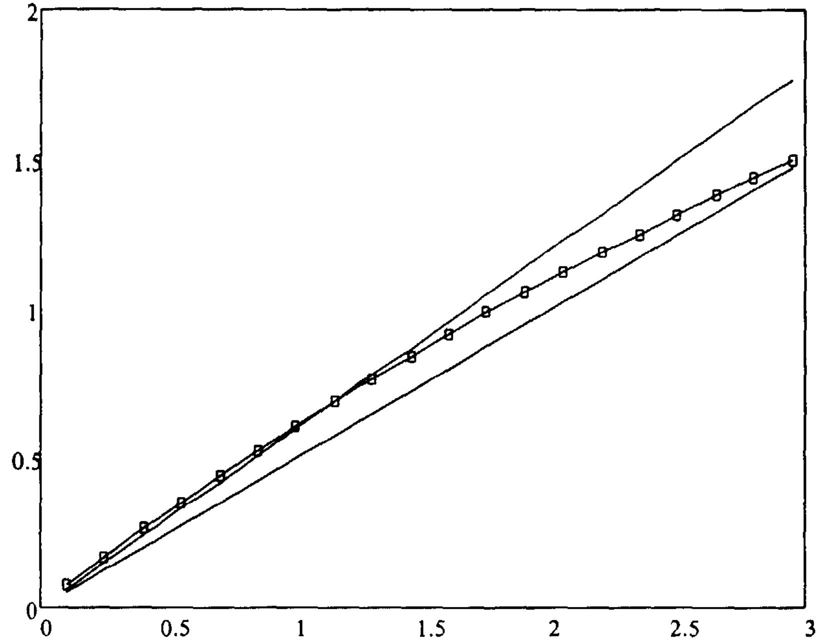


Figure 8: Structure Function for Interest Rate

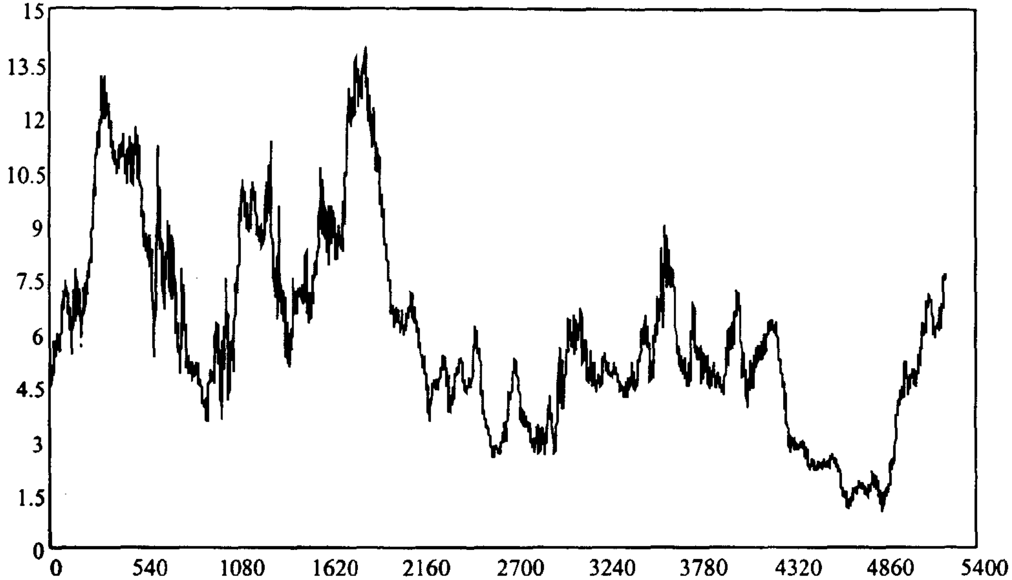


Figure 9: 5,000 simulated interest rates

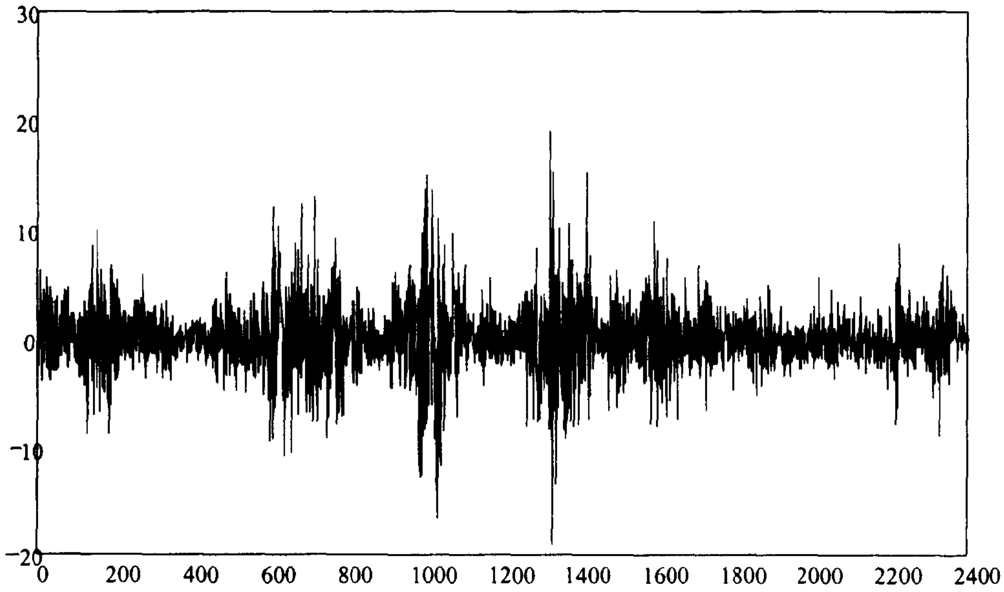


Figure 10: Logarithmic changes of Figure 9

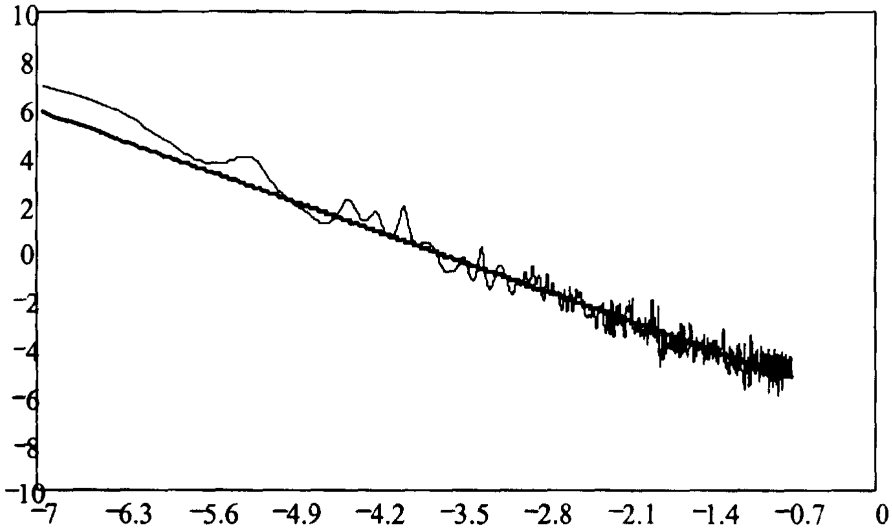


Figure 11: Power spectrum of simulated interest rate

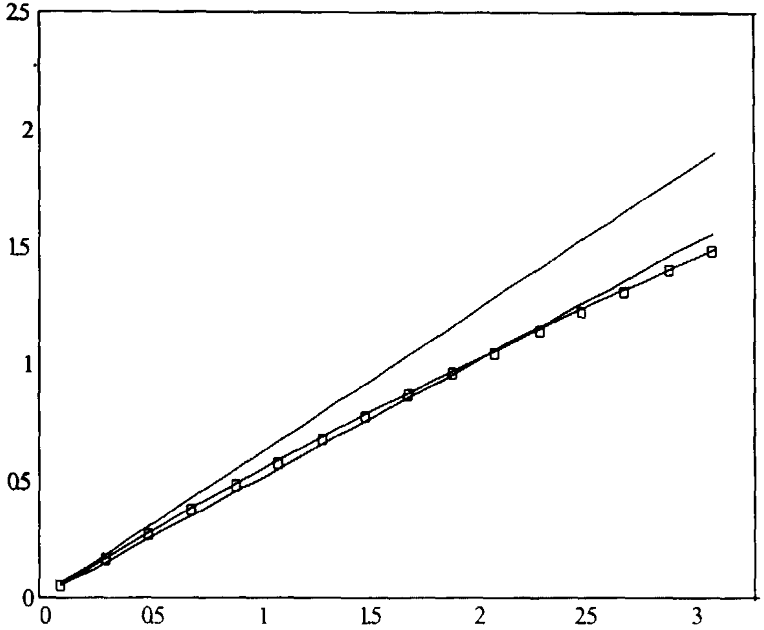


Figure 12: Structure Function for simulated interest rate

Let Me See: Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

"Human inside"

From the potential ad campaign

Dossier

Aleksey Popelyukhin is a Senior Vice-President of Technology with the Sam Sebe LLC and a Vice-President of Information Systems with the Commercial Risk Re in Stamford, Connecticut. He holds a Ph.D. in Mathematics and Mathematical Physics from Moscow University (1988).

His actuarially related achievements include:

- *Prize for the best 1997 article in the "Data Management discussion paper" program entitled "The Big Picture: Actuarial Process from the Data Management point of view" (1996)*
- *Creation and distribution of the popular actuarial utilities like Triangle Maker™ (1994) and Triangle Maker™ Pro (1997), Actuarial Toolchest™ (1998) and Enabler™ (1999)*
- *Design, development and coding of the 2nd and 3rd (current) generation of the very powerful and flexible actuarial software package called Affinity (1996)*
- *Promotion (through his papers and presentations) of his notions like Ideal Actuarial System and Data Quality Shield, and paradigms like object-oriented actuarial software and data-driven visualization.*

Aleksey is presently developing an integrated pricing/reserving/DFA computer system for reinsurance and also an action/adventure computer game tentatively called "Actuarial Judgement". Dr. Popelyukhin is an active member of several scientific societies and an author of almost 20 scientific publications.

Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

Abstract

No one would argue that there are limits on how much information a human being can perceive, process and comprehend. Even as advances in computer technology throw more and more data at actuaries, these limits stay the same. It is time to delegate to computers the very important task of *presentation of information.*

The article will try to demonstrate how existing *data-driven* technologies can help to evolve an *Ideal Actuarial System* from an actuarial tool into a company's *Alarm System.* Utilizing tools readily available to everyone who owns a contemporary Office Suite package, actuaries can present information in such a way that the effectiveness of *Corporate Decision Making, Data Error Detection* and suitability of *Actuarial Algorithms* will increase dramatically.

Actuarial results, properly combined, summarized and filtered **by importance**, may be arranged into a so-called *Digital Dashboard* that serves as a *portal* into the wealth of detailed actuarial information and the calculations behind it. This article itself can be considered as a portal that refers actuaries to the wealth of information on visualization techniques and data-driven technologies.

Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

Introduction

"Let's see..."

Graeae to Polyphemus

The Actuarial Process, like every analytical process, consists of three major stages:

- *Data Collection, Cleanup and Transformation*
- *Application of Algorithms and Methods*
- *Representation of Results*

All these stages require human interaction:

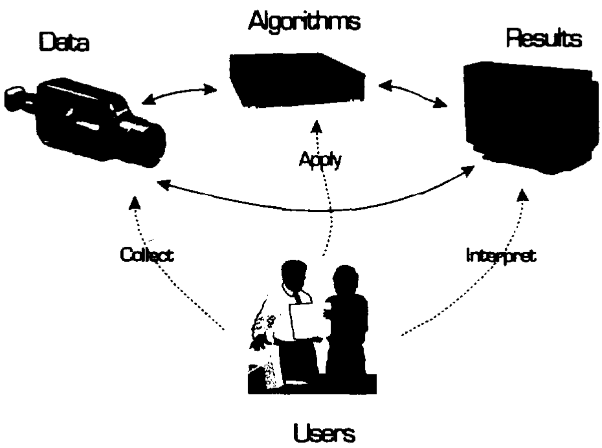


Figure 1

The amount of information to be processed is so vast, though, that there is no way to do it without computers.

But, computers are used strikingly differently throughout the process: in the first two stages we *serve* computers (feeding them with data and supplying them with actuarial parameters selection), and only in the last stage computers *serve* us (providing us with answers for making decisions). The irony here is that first two stages (*data transformation* and *application of algorithms*) have been automated to a fair degree, while *presentation of results* stage still remains the least computerized of all, and tools available for reporting and visualization are severely underused by actuaries.

This article is the first attempt to correct this situation.

Approaches

The amount of numeric information at all stages of the Actuarial Process is either exceedingly massive (raw data), overwhelming the recipient, or too small (summary) to adequately represent all the nuances and data patterns.

Human perception relies heavily on Short-Term Memory (STM) – a small "buffer" where external information is recognized and perceived (see [1]). Unfortunately, STM is limited in capacity – it can hold only 5-7 similar items at once. Another problem with STM is that new information replaces old (it's just a "buffer"), and new information is coming in continuously. Every change (in color, size or position) attracts a person's attention and changes the focus of his perception. Even in the best-case scenario, without any distractions, STM can hold information only 30 seconds or so[♦].

A few approaches seem to alleviate these limitations of human perception (and even exploit its weaknesses to attract a person's attention toward the important information): visualization, adaptive reporting and alarm systems. Indeed, "traditional" ways of displaying just myriad "boring" numbers in a spreadsheet are not adequate anymore. Without proper assistance, it is practically impossible to notice imperfections in the data; the inapplicability of a particular actuarial technique; or to pinpoint a claim, line or location that demonstrate unusual development.

The solution lies in augmenting standard report techniques with the information ***filtered by importance***. It means that only a few outstanding items with ***alarming*** behavior show up (or somehow get highlighted) in the report. The task of defining alarms and assigning levels of importance to actuarial results lies squarely on actuarial shoulders.

It is important to realize that nowadays, with the proliferation of Office Suites software, a wide variety of visualization tools is within the reach of every actuary. Almost every chapter of this article is illustrated by an example from an Office program. Equally important, one can safely assume that everybody understands the text of the BASIC[i] program. Coding in VBA has become a skill nearly as essential and vital as reading and writing.

♦ Conduct the following experiment: read a new telephone number digit-by-digit and then, (without attempting to repeat it, combine digits into groups or make associations and other mnemonic rules) after 30 seconds pass, try to dial it. Even without distraction it is practically impossible - that is probably why 411-type services repeat telephone numbers at least two times.

Adaptive Reporting

"Data is not necessarily Information."

Report Designer's Commandment

Every company almost certainly has established a fixed way of presenting the results of actuarial analysis. The overwhelming majority of these presentations are *static* reports with predefined content and layout: think of it as a list of reserves for 100 lines of business or a list of net present value of premium for 1000 treaties. There is nothing wrong with that way of presenting information, except that human perception cannot effectively span beyond 5-7 similar items. Nobody can guarantee that equal attention will be paid to every item in the long, monotonous report. To alleviate this problem, sometimes information is presented in a summarized form without important details. Either way, important information about the 68^{th} LOB or the 729^{th} treaty may escape the reader's attention.

The solution lies in the use of data-driven technologies to create *dynamic* or *adaptive* reports. Reports whose size, shape and format adapts to the data. Placing these reports in an interactive environment such as a spreadsheet allows the user to interact dynamically with the report (effectively creating a whole family of reports rather than a single one), shaping it to the level of detail that suits the user.

A partial list of data-driven implementations found in spreadsheets includes:

- *Filtering,*
- *Outlining,*
- *Sorting,*
- *Conditional Formatting and*
- *OLAP-enabled tools.*

Filtering

The simplest and most straightforward way to reduce the amount of information displayed in the report is Filtering. If information is organized as a list or a table in a spreadsheet and there is an easy way to define relevant subset, then Filtering fits the bill.

The AutoFilter feature of Microsoft Excel is a powerful and elegant implementation of Filtering. Accessible and customizable through either an interface or the VBA "macro" language, AutoFilter serves as an ideal Filtering tool*.

---

* Reporting and Visualization tools, including Filtering, are available in many products, not only in Excel. In fact, database products with built-in SQL language provide much more powerful and robust Filtering tool. However, these products may be outside the reach of many actuaries.

One can use this tool, for example, to limit the visible data table to a particular LOB[ii], location and Open/Closed status. Another interesting use of Filtering tools, and Excel's AutoFilter in particular, is checking for all distinct values in a list: AutoFilter's drop-down boxes display a sorted list of all distinct values present in a column. The fastest way to check if a huge spreadsheet populated with the data contains all requested LOB's, States or Policy Numbers is to initiate AutoFilter and click on the down-arrow in the corresponding column.

	A	B	C	D	E	F	G
1	Company	State	LOB	PaidInc	AccYear	EvalYear	NetLoss
112	ABC	CT	(All)	Paid	1988	1988	666,434
113	ABC	CT	(Top 10...)	Paid	1988	1989	1,143,933
114	ABC	CT	(Custom...)	Paid	1988	1990	1,342,429
115	ABC	CT	AL	Paid	1988	1991	1,432,131
116	ABC	CT	GL	Paid	1988	1992	1,525,594
117	ABC	CT	(Blanks)	Paid	1988	1993	1,625,430
118	ABC	CT	(NonBlanks)	Paid	1988	1994	1,682,298
119	ABC	CT	GL	Paid	1988	1995	1,702,690
120	ABC	CT	GL	Paid	1988	1996	1,766,216
121	ABC	CT	GL	Paid	1988	1997	1,730,972
122	ABC	CT	GL	Paid	1989	1989	704,188
123	ABC	CT	GL	Paid	1989	1990	1,272,152
124	ABC	CT	GL	Paid	1989	1991	1,464,204

Figure 2

Example 1. For an illustration of using AutoFilter for something less straightforward than plain-vanilla filtering (i.e., by LOB or Location), observe how to use it to filter losses in the 90[th] percentile of their Incurred Value:

```
Const colLOSS As Integer = 6       'Loss value is located in the column number colLOSS
Sub CreativeUseOfAutoFilter()
  Dim rRange As Range,nRows As Long
  ActiveSheet.Cells(1, 1).Select
  With Selection
    nRows = .CurrentRegion.Rows.Count
    Set rRange = .Offset(1, colLOSS - 1).Resize(nRows - 1, 1)
    .AutoFilter
    .AutoFilter Field:=colLOSS, Criteria1:=">" & Application.Percentile(rRange, 0.9)
  End With
End Sub
```

Filtering is a fast and effective way to cut down the amount of data displayed. However, if a filtered subset is still too large, or there is a need to see different levels of detail for different groups of data, or the user has to see different aggregate values (subtotals, averages), then Outlining or Pivot Tables would be a better choice of tools.

Outlining

Outlining is a hierarchical representation of data organized into a list or table, with the ability to hide or display details of all or selected groups on any level of the hierarchy. Every user of Windows Explorer or any other File Directory tool is familiar with the notion of Outlining. Excel's implementation of Outline allows both horizontal (rows/records) and vertical (columns/fields) outlining. Along with the ability to display detailed records, Excel's Outline supports aggregate functions such as sum, count, average, standard deviation, etc. This capability may become handy in situations when different members of a hierarchy are treated at different levels of detail.

Example 2. A screenshot below illustrates Outline's advantages over AutoFilter: it allows the display of different levels of detail for different LOB's on the same screen. While WC information is shown up to the State and Accident Year levels, GL and AL are shown only as aggregates without unnecessary details.

	A	B	C	D	E
1	Reinsured	LOB	State	AccYear	UltNetLoss
2	ABC	WC	NY	1996	1,712,201
3	ABC	WC	NY	1997	1,730,918
4			ABC WC NY Total		3,443,119
5	ABC	WC	CT	1996	1,944,502
6	ABC	WC	CT	1997	1,975,489
7			ABC WC CT Total		3,919,991
8	ABC	WC	NJ	1996	2,172,041
9	ABC	WC	NJ	1997	2,227,708
10			ABC WC NJ Total		4,399,750
11		ABC WC Total			11,762,860
14		ABC GL Total			14,245,270
17		ABC AL Total			7,249,632
18	ABC Total				33,257,762
48	XYZ Total				32,809,931
49	Grand Total				66,067,693

Figure 3

Example 3. Microsoft Access 2000's tables with properly defined "master-detail" relationships can display records in an Outline fashion too, making table navigation as simple and intuitive as browsing a directory tree in Windows Explorer.

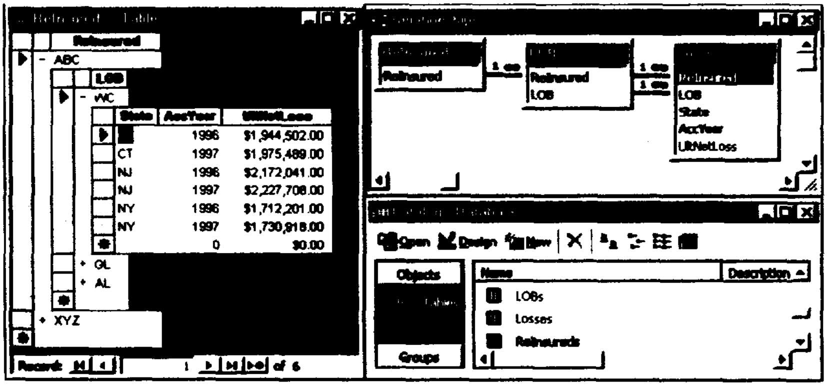


Figure 4

Sorting

Sorting is a powerful technique that brings the most important records to the top of the display or, in the case of printed report, to the first page. Sorting does not reduce the amount of data displayed, but it assures that the first several records get more attention before the report reader gets tired. Consequently, the main skill of using Sorting is in defining what constitutes ***importance***. An ability to rank information in accordance to its significance, and to identify which actuarial measurements affect decision-making ***the most*** is a yardstick that separates great report designers from mere mortals.

"Combined Loss Ratios," "Net Profitability" and "Reserve Adequacy" – all these measurements may serve as actuarial significance indicators. Both more sophisticated ones like "Percent Change of the Current Estimate of the Net Ultimate Loss Hindsight Estimated" or simpler ones like "Time since the Latest Loss Run" help to sort data and generate useful actuarial information.

Sort | 20%

Aleksey Popelyukhin: Enter desired % and press "Sort" button to resort

UW Year	Policy	Premium	Aggr Limit	Aggr Ret	Prob to lose 20% Prem
1998	NWA-00019	$ 2,944,691	$ 5,700,000	$ -	0.22
2000	NWA-00003	$ 913,502	$ 2,500,000	$ -	0.21
2000	NWA-00007	$ 6,736,411	$ 8,100,000	$ -	0.21
1998	NWA-00011	$ 4,226,719	$ 9,900,000	$ -	0.21
2000	NWA-00030	$ 1,224,858	$ 2,100,000	$ -	0.19
2000	NWA-00006	$ 7,994,304	$ 24,600,000	$ -	0.18
1998	NWA-00012	$ 1,594,191	$ 8,500,000	$ 2,800,000	0.18
1998	NWA-00002	$ 5,818,746	$ 29,100,000	$ 9,700,000	0.17
2000	NWA-00015	$ 9,135,813	$ 31,800,000	$ -	0.16
1998	NWA-00010	$ 4,837,832	$ 5,200,000	$ -	0.16
2000	NWA-00017	$ 8,455,004	$ 45,200,000	$15,000,000	0.16
2000	NWA-00001	$ 5,817,872	$ 33,200,000	$11,000,000	0.16
1998	NWA-00023	$ 8,313,738	$ 13,800,000	$ -	0.15
1998	NWA-00032	$ 254,889	$ 900,000	$ -	0.15
1998	NWA-00024	$ 2,723,106	$ 10,500,000	$ -	0.14
1999	NWA-00028	$ 6,413,104	$ 23,100,000	$ -	0.13

Figure 5

Example 4. In a situation where a new significance indicator has to be added or modified, there is the need to copy an analytical expression into all the cells adjacent to the existing data table. The code below helps in these situations:

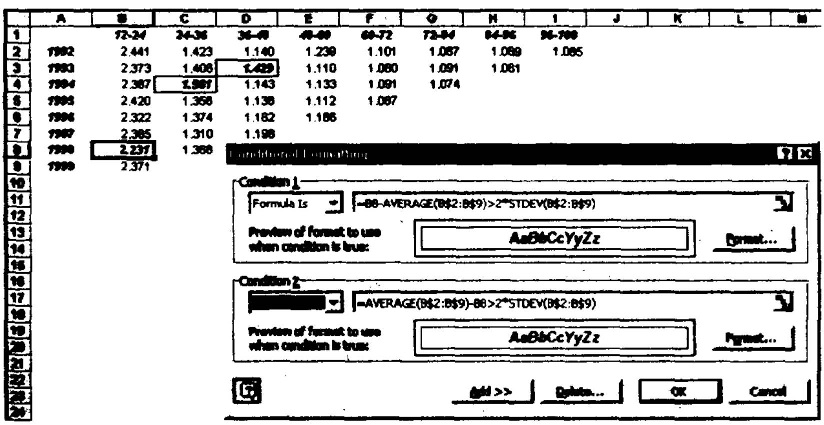
```
Sub FillAdjacentColumn()
   With ActiveCell
      .Resize(.CurrentRegion.Rows.Count + .CurrentRegion.Row - .Row, 1).Formula _
        = .Formula
   End With
End Sub
```

Conditional Formatting

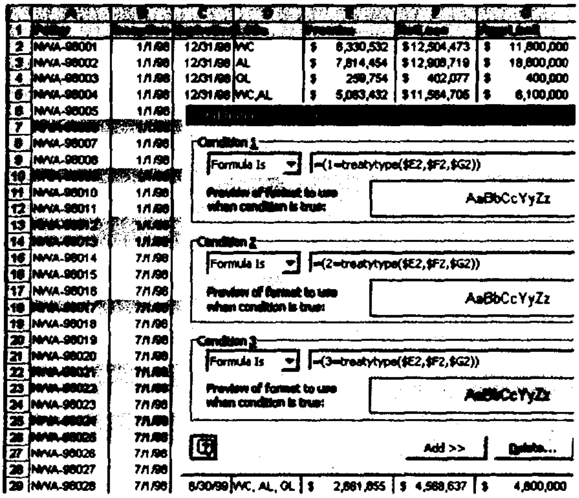
Conditional Formatting is a feature of Microsoft Excel that allows users to define the font, color, border and background pattern of a cell as a function of the values in other cells. When values in the referenced cells change so does the conditional format. Despite some limitations (currently, Excel currently supports up to 3 variable formats per cell), this feature opens unprecedented creative possibilities for report designers. Combined with other reporting techniques like Filtering, Sorting and Pivot Tables, Conditional Formatting is indispensable for attracting the report reader's attention to the most crucial information. Due to its dynamic nature, Conditional Formatting can serve as a building block for an actuarial Alarm System (a cell automatically becomes red when an actuarially significant value becomes too high or too low).

Given that the format condition's formula can be any expression that uses user-defined functions along with built-in ones, the use of Conditional Formatting is limited only by one's imagination: it's use can range from data-error detection to Thomas Mack-like assumption testing to statistical outlier warnings (see [2]).

<u>Example 5</u>. A powerful yet simple application of Conditional Formatting would be highlighting statistical outliers. For example, it would be convenient to see values in a triangle of age-to-age factors that are too far (more than 2 or 3 standard deviations) from the column's average. Outliers like these frequently trigger an additional investigation of data that usually produces interesting results:



<u>Example 6</u>. As an example of the user-defined format condition formula, one can use an algorithm that assigns different types to the treaties based on the relationships among parameters like Premium, Ultimate Loss, Aggregate Limit and others. Every treaty's record in this kind of report can be formatted in accordance with assigned type. And as the estimate of the ultimate net loss changes, the type of the treaty (potentially) changes, and so does the formatting.



Conditional Formatting used in combination with other reporting techniques is especially powerful. Sorting a list of records by one criterion (e.g., Net Profitability) and Conditionally Formatting them according to another criterion (e.g., treaties with a particular insurance company can create quite an impressive display and produce an easy-to-comprehend overall picture.

OLAP

OLAP stands for Online Analytical Processing and there are many tools from numerous vendors that provide OLAP functionality. However, in order to demonstrate the accessibility of these tools, only one particular implementation of OLAP will be considered – Microsoft Excel's Pivot Tables.

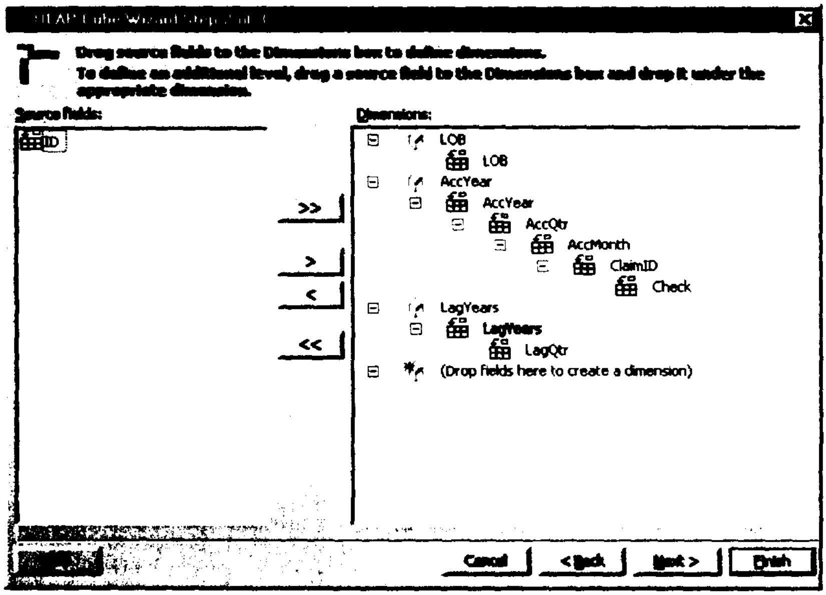
A Pivot Table is essentially a mechanism for displaying multidimensional information in 2-D. Before data can be used in a Pivot Table, it is converted into multi-D array called an OLAP cube, where the table's fields become dimensions. That is, a $1,000,000 1999 WC California paid loss as of 6/30/2000, becomes an element (value, measure) of the OLAP cube with dimensions such as State, LOB, Accident Year, Evaluation Date stored at the intersection of the "CA," "WC," "1999," and "6/30/2000" members of these respective dimensions). There are only two meaningful ways to display multi-D info in 2-D: cross-section and projection with aggregation – and both ways are implemented in Excel's Pivot Tables. When used to display a cross-section of the data, a Pivot Table serves as a Filtering tool; when used to display aggregations (subtotals, averages, etc.), Pivot Table serves as an Outline*. Pivot Tables also support Sorting and, to some degree, Conditional Formatting. So, in effect, this tool incorporates all the adaptive reporting techniques described above, and as such it should serve as a **preferred choice** for any report designer. In addition to that, Pivot Tables are capable of using external data sources and, consequently, are subject to fewer limitations than Excel's Sort, Filter and Outline.

* Bear in mind, that for true Outlining, it is necessary to support hierarchies. Starting with Excel 2000, Pivot Tables' dimensions do support hierarchies. See [3] for examples of actuarial hierarchies.

The total flexibility of Pivot Tables may cause some problems for actuaries.

First, it is unclear which dimensions to choose for display and which ones for aggregation (or cross-section) to get actuarially meaningful results. Also, actuaries should define additional (calculated) fields with some kind of "actuarial significance" indicator, which can be later used in Sorting, Filtering or Conditional Formatting.

Second, unlike other professions where creating a Pivot Table is a destination – a final act of the analytical process – actuaries frequently use aggregated data as a starting point of their analysis (see [4]). If created properly, a Pivot Table can serve as a convenient storage for actuarial triangles with selective drill-down capabilities. One can create an OLAP Cube hierarchy in such a way that any suspicious element of the triangle can be drilled-down for details up to individual Claims and even individual payments level:



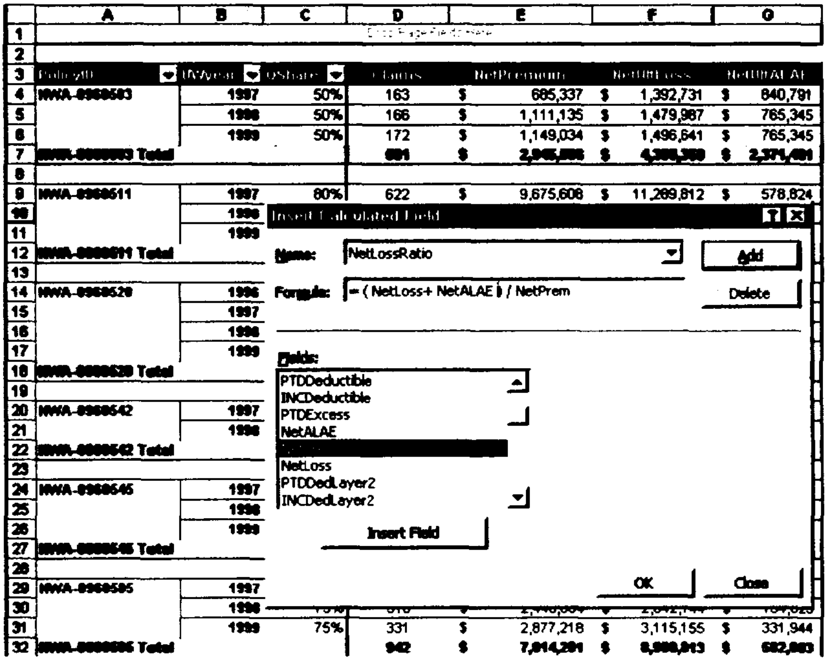
	A	B	C	D	E	F	G	H	I
1	LOB	WC							
2									
3	Sum Of Loss					LagYears			
4	AccYear	AccQtr	AccMonth	ClaimID	Check	1	2	3	4
5	1996					$2,166,733	$411,335	$408,024	$366,587
6	1997					$1,094,331	$381,124	$117,174	
7	1998					$1,225,399	$67,132		
8		2				$39,240	$67,132		
9		3				$19,429			
10		4				$1,166,730			
11			11			$70,027			
12			12			$1,096,702			
13				C-00240		$1,070,676			
14					1023	$54,176			
15					1036	$12,514			
16					1694	$14,577			
17					1784	$989,409			
18				C-00501		$26,027			
19	1999					$2,047,854			

Figure 6

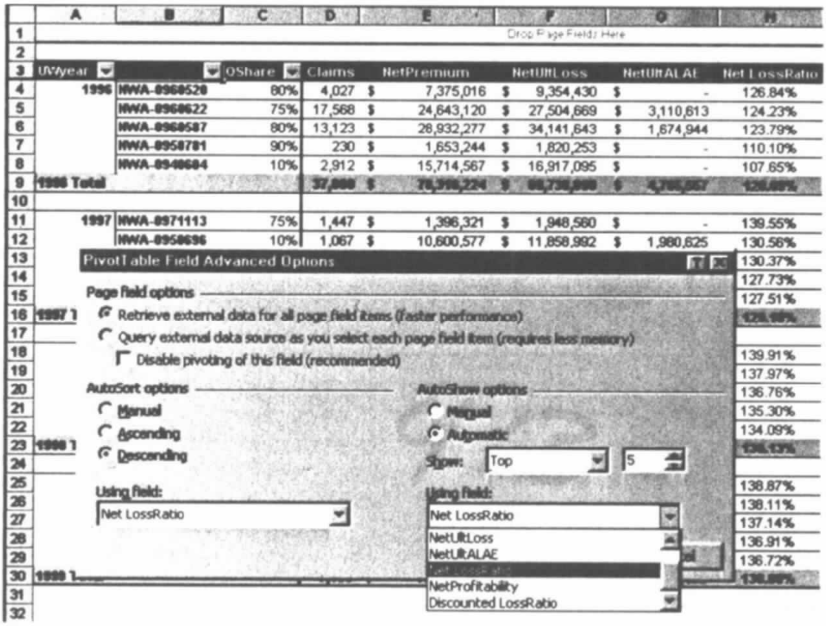
The problem is that with an unpredictable shape/size of the Pivot Table it is hard to incorporate its content into subsequent calculations. One workaround is to use the GETPIVOTDATA spreadsheet function, while another is to use the Pivot Tables' Calculated Fields – an ability to add fields/dimensions to the Pivot Table that are calculated on the fly.

> Example 7. Sometimes, a Calculated Field is the only mechanism to add new dimensions correctly. As an example, consider the Loss Ratio field. A Loss Ratio like any ratio is a nonlinear operation and, consequently cannot be summarized properly: a sum of ratios *is practically never*[*] a ratio of the sums. That is where Excel's Pivot Tables make a clear distinction between input fields (for subtotals, the *sums of ratios* are calculated) and Calculated Fields (for subtotals, the *ratios of sums* are calculated).

[*] Unless the ratios have the same absolute values and different signs.



Example 8. Once a (calculated on the fly) Loss Ratio field is added to the Pivot Table, one can use it for Sorting and Filtering. A screenshot below illustrates 3 important features of Pivot Tables simultaneously. First, by the simple dragging of the field label, one can rearrange the Pivot Table from a "Policies by Underwriting Years" to an "Underwriting Years by Policies" view. Second, one can Sort a field by the results of on-the-fly calculations: in our case we sorted Policies by Loss Ratios in descending order. And, third, one can Filter by the results of on-the-fly calculations; in our case we chose to display just the 5 worst policies per underwriting year based on the Loss Ratio indicator. Note that we could choose any indicator (like Net Profitability or Discounted Loss Ratio) that is available in the Pivot Table as an input field or a dynamically calculated field. As one can see, the impressive reporting tools are all there; the quest is on for actuarial indicators.



Visualization

Seeing is believing

Popular belief

Visualization (see [5]) is the process of exploring, transforming and viewing data as images to gain understanding and insight into the data. Studies in human perception, computer graphics, imaging, numerical analysis, statistical methods and data analysis have helped to bring visualization to the forefront. Images have unparalleled power to convey information and ideas. Informally, visualization is the transformation of data or information into pictures… it engages the primary human sensory apparatus, *vision*, as well as the processing power of the human mind. The result is an effective medium for communicating complex and/or voluminous information.

As the amount of data overwhelms the ability of the human to assimilate and understand it, there is no escape from visualization. So actuaries have to develop conventions for representation of their data and results.

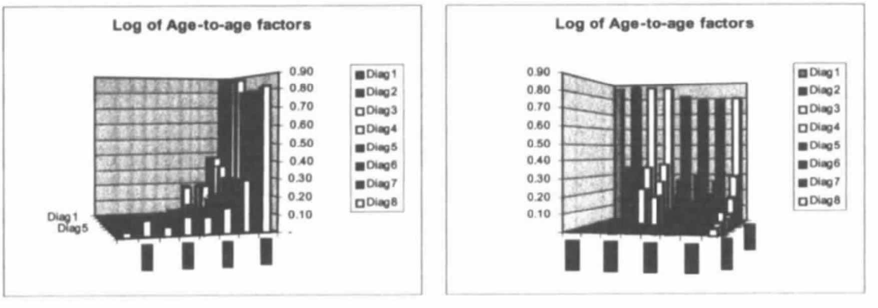
There exist a multitude of visualization approaches: mapping scalars to colors, contouring (isosurfaces), glyphs (arrows of different color, length, direction), warping (display of different stages in the motion), displacement plots, time animations, streamlines (particle traces) and tensor algorithms.

There are unavoidable problems with multidimensional visualization: projection (only two or three dimensions are available) and understanding (we as humans do not easily comprehend more than three dimensions or three dimensions plus time animation). Projections can be implemented if the three most important dimensions are identified in such a way that the remaining dimensions can be ignored. Once again, it is an actuarial task to choose these dimensions.

Charts

This paper is too short to discuss all the possible uses of charts and diagrams in the actuarial process. A great number of wonderful books (see [6]-[8]) explain which type of chart to use in which situation: a line chart for displaying increases, bar for shares, pedestal for ranks, Gantt for schedules, etc... However, it is up to actuaries to decide how to display triangular data.

> Example 9.By examining the 3-D chart representing logarithms of age-to-age factors, one can formulate a hypothesis about changes in calendar year trends (arguably not immediately evident from looking at the raw data): the last 4 years have different trend than the rest of the triangle. Rotating the graph for the look from another angle allows one to confirm or discard that hypothesis. The final check comes from the algorithm described, for example, in [9].



	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108
1990	0.892	0.353	0.241	0.214	0.097	0.083	0.067	0.081
1991	0.864	0.341	0.257	0.118	0.058	0.087	0.059	
1992	0.870	0.367	0.175	0.125	0.087	0.071		
1993	0.884	0.304	0.130	0.106	0.084			
1994	0.801	0.317	0.150	0.154				
1995	0.802	0.270	0.181					
1996	0.797	0.313						
1997	0.797							

Figure 8

For almost 86 years since 1914, CAS members have not yet agreed on the standard graphical representation of a triangle (one of the most basic actuarial notions). The author firmly believes that properly displayed triangles may reveal some important trends that are not evident in numerical representations.

Not only does visualization inspire useful hypotheses for actuarial analysis: sometimes it's the only way to deal with the data♣. Developments in new areas of actuarial science present new challenges in demonstrating findings. DFA[iii] that deals with many less traditional notions, such as scenarios and strategies, is a good example of such a challenge.

Example 10. The majority of actuarial information contains a location code associated with the values. Legislation requirements, types of coverage, rates, exposures and loss performance differ from region to region. Geo-coding swiftly emerges as one of the hottest actuarial applications. Yet it is hard to imagine how one can notice trends and dependencies of geographically related data without visualization. Microsoft MapPoint – ideologically, an integral part of Microsoft Office – provides the means for precisely that type of visualization. For example, a map of WC Ultimate development factors by State based on NCCI data (see [10]) is presented below.

Figure 9

Animation

If a picture is worth a thousand words, then an animation is worth a thousand pictures.

Animation is the best way to exploit an aspect of the human psyche called "selective attention" – people readily react (by shifting the focus of their attention) to any movement, including change in color, size or position. Animation is suitable for visualization of the range of uncertainty and/or development – two of the most important actuarial phenomena. While not a standard feature of a plain-vanilla spreadsheet, animation is nevertheless quite within reach of every Microsoft Excel user.

♣ So-called "quarterly" triangles, sometimes studied by actuaries, can easily reach size of 60x60 or more, which makes them impractical for examining by traditional means (in a spreadsheet), yet visualization techniques would shine in this circumstance.

Example 11. Below is the code for conceptual animation:

```
Sub AnimationInExcel()
Dim i As Long
   Call SetupAnimation     'see Appendix 2
   Application.Iteration = True
   Application.MaxIterations = 1000
   Application.MaxChange = 0.1
   For i = 1 To 1000
      Calculate
   Next i
End Sub                        'do not forget to restore original Calculation Mode!!!
```

Taking a tip from the computer games industry, animation can be **effectively** used for visualization of the simulation process (see [11]). Indeed, with the growing importance of DFA and other non-analytical modeling techniques, simulations are steadily becoming a technique of choice for the majority of actuaries. Currently, however, all intermediate steps used in simulation are hidden from the user – such a wealth of information is, essentially, discarded. Use of animation may prevent this "waste" of intermediate calculations. Animated display of simulation's steps may help the user to visualize the dynamics of the simulated process or appreciate the range of uncertainty in simulated scenarios.

Interactive selection

Another application of visual technologies is the interactive selection of actuarial parameters. For example, selecting parameters by moving points directly on the graph of a development pattern appears to be much more intuitive and convenient than typing numbers into spreadsheet cells.

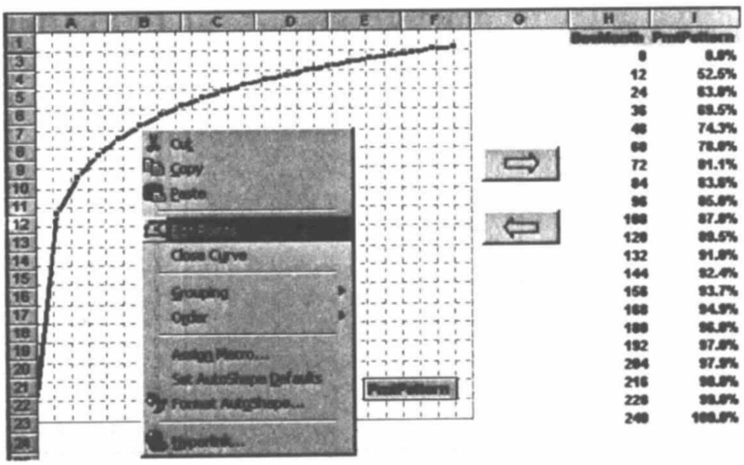


Figure 10

This functionality – a two-way link between the graphical display of numbers and their values in a spreadsheet – is not yet as user-friendly as other Office tools. But given that this feature is available in numerous other applications, it is only a matter of time until Visual Editing becomes an equal member of the Office tools roster and actuaries are able to incorporate interactive graphical manipulation of numbers into their spreadsheets.

Example 12. Even though "Visual Chart Editing" is not a standard Office feature, with some amount of VBA programming, it is possible to establish a two-way link between numbers and shapes in Excel (below is the code behind the buttons from Figure 10):

```
Sub VisualizationFromShapeToSpreadsheet()
   Dim cell As Range, node As ShapeNode, n As Integer
   Set cell = ActiveCell
   ActiveSheet.Range("Coordinates").CurrentRegion.Clear
   ActiveSheet.Shapes("InteractiveSelect").Select
   n = 0
   For Each node In ActiveSheet.Shapes("InteractiveSelect").Nodes
         n = n + 1
         ActiveSheet.Range("Coordinates").Cells(n, 1) = node.Points(1, 1)
         ActiveSheet.Range("Coordinates").Cells(n, 2) = (210 - node.Points(1, 2)) / 200
   Next node
   cell.Select
End Sub

Sub VisualizationFromSpreadsheetToShape()
   Dim c As Range, cell As Range
   Set cell = ActiveCell
   On Error Resume Next
   ActiveSheet.Shapes("InteractiveSelect").Delete
   With ActiveSheet.Shapes.BuildFreeform(msoEditingAuto, 0, 0)
      For Each c In ActiveSheet.Range("Coordinates").CurrentRegion.Resize(, 1)
         .AddNodes msoSegmentLine, msoEditingAuto, c.Value, 210 - 200 * c.Cells(1, 2).Value
      Next c
      .ConvertToShape.Select
   End With
   Selection.Name = "InteractiveSelect"
   ActiveSheet.Shapes("InteractiveSelect").Nodes.Delete 1
   Selection.Placement = xlFreeFloating
   cell.Select
End Sub
```

Alarm Systems

It's the eleventh hour – do you know where your reserves are?

Actuarial proverb

Data Quality

Companies that build data warehouses and clean up their data soon realize that the majority of their data comes from external sources (TPA's[iv], industry bodies, self-insureds), which are neither clean nor in a single format. It is time to combine efforts and make sure that every source can supply high quality data in a timely manner.

There are some recommendations on data quality procedures by IDMA[v] (see [12], [13]) and data elements' definitions by ISO[vi] and NCCI[vii] (see [10], [14]), but they are not part of everyday life in every data collection entity. In fact, a study of more than 40 TPA's (see [15]) showed that practically every one of them has failed even the most primitive data quality checks.

Example 13. An Alarm System that is worth its while should trigger some action when a problem is found. Painting some cells in a spreadsheet is a good example of such an action, but automatically sending an e-mail with the description of the problem would be much more effective. The code below continues Example 5: first, it checks a triangle of age-to-age factors for outliers and then it sends an e-mail to the System Administrator with the addresses of all problematic cells:

```
Sub AlarmEvent()
  Dim c As Range, sAlarm As String
  Dim otlApp As Outlook.Application, eMail As Object
  For Each c In Range("Tri").SpecialCells(xlCellTypeSameFormatConditions)
    c.Select
    Range("Temp").Formula = c.FormatConditions(1).Formula1
    If Not Application.IsError(Range("Temp").Value) Then
      If Range("Temp").Value Then
        sAlarm = sAlarm & "Problem at :" & c.Address & Chr(13)
      End If
    End If
  Next c
  If sAlarm & "" <> "" Then                    'if there is any problem
    Set otlApp = New Outlook.Application       'launch Outlook
    Set eMail = otlApp.CreateItem(olMailItem)  'and create e-Mail message
    With eMail
      .To = "samsebe@consultant.com"
      .Subject = "ALARM from " & ActiveWorkbook.Name & "!" & ActiveSheet.Name
      .Body = sAlarm
      .Send                                    'send e-Mail
    End With
    Set eMail = Nothing
    otlApp.Quit
    Set otlApp = Nothing
  End If
End Sub
```

Data Quality can be tested both on the detail level and on pre-aggregated levels (see [15]). In both cases, reporting techniques like Filtering, Sorting and Conditional Formatting may help attract attention to the problem (the same applies to visualization techniques, which can help to pinpoint a problem). One can calculate changes in case reserves and sort claims in descending order by that field to bring the largest outstanding claims on top. Or, using conditional formatting, one can highlight outliers among age-to-age factors (see Example 5).

Algorithms Applicability

Closely related to Data Quality tests on the pre-aggregated (as opposed to detailed) level is actuarial assumption testing (see [15]). Indeed, a monotonically increasing number of claims can be both a data quality test and a requirement for the applicability of the Berquist-Sherman algorithm. The same for the assumption of lognormality in ICRFS[viii], which coincides with the check that requires incremental gross payments to be positive. An Alarm System may warn users about *Thomas Mack*-style test (see [2], [16]) failures.

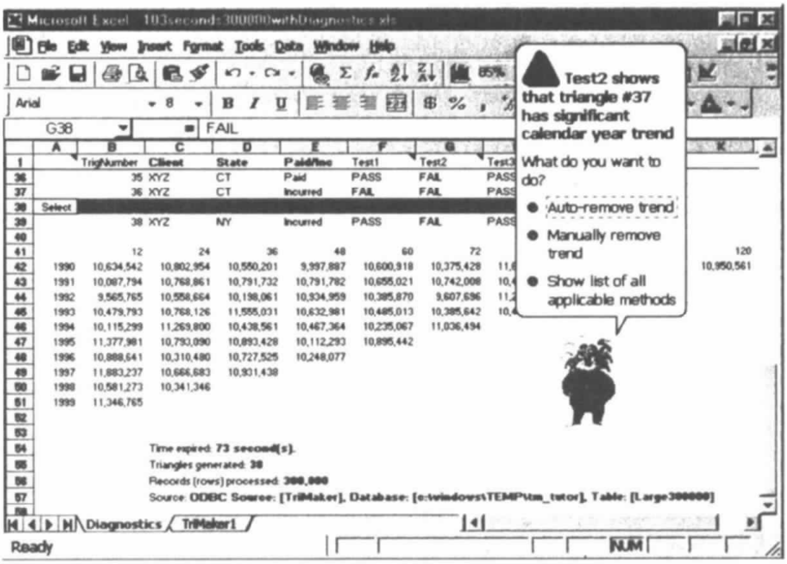


Figure 11

Digital Dashboard

Digital Dashboard is Microsoft's name for a portal that consolidates the most important personal, professional, corporate and external information with an immediate access to analytical and collaborative tools. In a single view, the user can see charts, Alarm messages, Pivot Tables, calendars, etc. Thus, Digital Dashboard looks like an obvious place for all important reports and alarms. Dashboard's space limitations re-emphasize the necessity of smart and space-conscious reporting techniques: Dashboard's start screen is the place for the most important information presented in the most concise way.

While every reporting and visualization technique described in this paper is powerful and effective, it is their combinations (Filtering + Sorting, Pivot Tables + Conditional Formatting, etc + etc…) that convert a flood of data into truly useful and indispensable information. Digital Dashboard – a "combination of combinations" of reporting tools - is just a very logical extension of the mechanisms that make this information immediately available and accessible. By the same token, Digital Dashboard is a most natural interface for an Alarm System. Not only can it display all types of alarms in a single location, but also – thanks to its portal capabilities – it can provide links to detailed information that triggered an alarm.

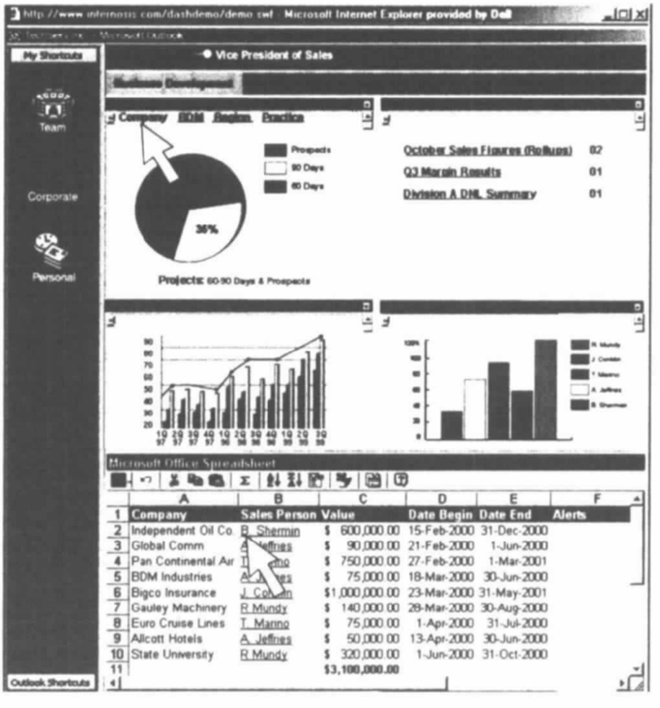


Figure 12

With the proliferation of the Internet, portal interfaces have become very popular: an ability to organize a wealth of information into a concise and focused display is very appealing. In fact, this article itself is organized as a portal into a wealth of information on reporting and visualization techniques: it is just a concentrated extract of the most important facts about tools available to actuaries, and as such it serves as a starting point for further research.

Actuarial Significance

"The whole system makes me feel... insignificant."

Ant Z.

Actuaries have a lot of work to do before any of the aforementioned techniques can be used to generate useful reports and visualizations.

For adaptive reporting, actuaries have to decide which measurements are of actuarial significance, so reports can be filtered or sorted accordingly. Every step of the actuarial process requires different significance indicators. For the *data preparation* step, actuaries should find measurements that will catch severe data errors that may considerably affect the consequent application of actuarial methods. For the *results presentation* step, actuaries should define (and calculate during the *application of algorithms* step) indicators that will aid decision-making. That will help to concentrate the attention of the report readers on important issues and will energize and strengthen the decision-making process.

Another important area that needs assistance from actuaries is in Alarm Systems. Nobody likes false alarms. It is actuary's job to come up with and fine-tune alarm definitions, to determine which combination of circumstances should trigger an actuarial alarm and attract immediate attention.

Selection of these *most important* variables depends on the available data and the goal of the display, and is clearly an actuarial task. It would make sense for actuaries to develop conventions that cover most situations. Unfortunately, to the best of the author's knowledge, this work has not even started.

Conclusion

"The gods help them that help themselves."

Aesop

A list of easily accessible presentation techniques with examples of their uses should help actuaries to realize what tools are available for Actuarial Reporting. But it is up to actuaries to express themselves using these tools. Indeed, reporting techniques described in this article are so flexible it does not make sense to use a limited number of pre-designed "canned" reports anymore. In addition to that, reporting tools are incredibly interactive – they were designed in order to give the end user (i.e., actuary) report-creation power. And they are so easy to use – it is a sin not to use them.

The importance of presentation skills is severely underestimated by actuaries: it is conceivable that quite a few companies would still remain solvent if actuaries-in-charge could **convincingly present** results of their analysis. If actuaries expect computers to be effective helpers in reporting and visualization tasks, they have to define "actuarially significant" information and learn how to present it in the most "attention-grabbing" way. That would insure that actuarial analysis is indeed used as a solid foundation for the company's decision-making process.

Acknowledgements

"Thank you very much."

Jim Carrey as Andy Kaufman as Latka Gravas

Many thanks to Leigh Walker who helped to make this paper a better presentation of the author's ideas: he issued an *alarm* every time he saw a grammatical error or an unclear passage. The author is eternally grateful to Boris Privman, FCAS who greatly affected his *views* on the actuarial profession. And without the understanding and support of the author's family, this article would have not been *presentable* at all.

Stamford, 2000

Appendix 0

Readers of the Acrobat version of this paper (downloaded from the www.casact.org website) can copy code snippets from the text and paste them into Excel's VBA Editor as described below.

Appendix 1

To run subroutines described throughout the article:

- *launch Excel, preferably version 97 or better*
- *press Alt-F11 to start VBA© Editor*
- *find a project with the name of your current spreadsheet or, if you prefer, Personal.xls*
- *right-click on the project and choose Insert/Module*
- *type or copy/paste the desired fragment of the code into the Module window*
- *launch code from the Excel: press Alt-F8, select macro name and press Run*

Appendix 2

```
Sub SetupAnimation()
Dim sName As String
  With ActiveSheet
    .Cells(1, 2).Formula = "T"
    .Cells(1, 3).Formula = "X"
    .Cells(2, 2).Formula = "12"
    .Cells(3, 2).Formula = "24"
    .Cells(10, 3).Formula = "1"
    .Cells(1, 1).FormulaR1C1 = "=RC+0.01"
    Range("B2:B3").Select
    Selection.AutoFill Destination:=Range("B2:B10"), Type:=xlFillDefault
    Range("C2:C9").Select
    Selection.FormulaR1C1 = "=1-EXP(-1/R1C1/(1-RC[-1]/R10C2)/(1-RC[-1]/R10C2))"
  End With
  sName = ActiveSheet.Name
  Charts.Add
  ActiveChart.ChartType = xlXYScatterSmooth
  ActiveChart.SetSourceData Source:=Sheets(sName).Range("B1:C10")
  ActiveChart.Location Where:=xlLocationAsObject, Name:=sName
  With ActiveChart.Axes(xlValue)
    .MinimumScale = 0
    .MaximumScale = 1.2
    .MinorUnit = 0.04
    .MajorUnit = 0.2
    .CrossesAt = 0
  End With
End Sub
```

Bibliography

(in the order of reference)

[1] Theo Mandel. *The elements of User interface Design.* Wiley Computer Publishing, 1997

[2] Thomas Mack, *Measuring the Variability of Chain Ladder Reserve Estimates.* CAS, 1993

[3] Aleksey S. Popelyukhin. *On Hierarchy of Actuarial Objects: Data Processing from the Actuarial Point of View.* CAS Forum, Spring 1999

[4] Aleksey S. Popelyukhin. *The Big Picture: Actuarial Process from the Data Processing Point of View.* Library of Congress, 1996

[5] Will Schroeder, Ken Martin, Bill Lorensen. *Visualiazion Toolkit.* Prentice Hall, 1998

[6] Gene Zelazny. *Say it with Charts: The Executive's Guide to Visual Communication.* McGraw-Hill, 1996

[7] Robert L. Harris. *Information Graphics: A Comprehensive Illustrated Reference.* Management Graphics, 1997

[8] Tufte Edward R. *Visual Exlplanations.* Graphics Press, 1997

[9] Ben Zehnwirth, *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital*

[10] *NCCI Annual Statistical Bulletin, 1998 Edition.* NCCI, 1999

[11] Aleksey S. Popelyukhin. *SimActuary: The Game We Can Play.* Submitted to CAS Forum, Spring 2003

[12] *Insurance Data Quality.* Marr, Richard, ed. IDMA, 1995

[13] *Data Quality Certification Model (Framework and Guidelines) for Insurance Data Management.* IDMA, ISO, 1995

[14] *Standards and Guidelines for Data Elements.* ISO, 1981

[15] Aleksey S. Popelyukhin. *Watch your TPA: A Practical Introduction to Actuarial Data Quality Management.* CAS Forum, Spring 1999

[16] Gary G. Venter, *Checking Assumptions of Age-to-Age Factors.* CLRS, 1994

Acronyms

---

[i] **BASIC - Beginner's All-purpose Symbolic Instruction Code**: one of the earliest and simplest high-level programming languages – still a very popular choice among educators.
[ii] **LOB – Line of Business**: here, a type of insurance coverage like Workers' Compensation or Professional Liability.
[iii] **DFA – Dynamic Financial Analysis**: a process for analyzing the financial condition of an insurance entity.
[iv] **TPA – Third Party Administrator**: a company in the business of handling day-to-day activities and/or providing services on insurance claims. Consequently, TPA is a primary source of actuarial data. See [14].
[v] **IDMA - Insurance Data Management Association**: an independent nonprofit association dedicated to increasing the level of professionalism in insurance data management. http://www.ins-data-mgmt.org .
[vi] **ISO – Insurance Services Office**, Inc.: leading supplier of statistical, actuarial, underwriting, and claims information. http://www.iso.com .
[vii] **NCCI – National Council on Compensation Insurance**, Inc.: a value-added collector, manager, and distributor of information related to workers' compensation insurance. http://www2.ncci.com/ncciweb .
[viii] **ICRFS – Interactive Claims Reserving Forecasting System**: commercially available statistical modeling framework from Insureware. http://www.insureware.com .
[ix] **VBA – Visual BASIC for Applications**: version of BASIC language embedded into host application (i.e. Excel) with the access to host's objects - a better "macro language". http://www.microsoft.com .

Materiality and ASOP No. 36: Considerations for the Practicing Actuary

CAS Committee on
Valuation, Finance, and Investments

November 17, 2000

To Actuaries Preparing Statements of Actuarial Opinion Regarding Property/Casualty Loss and Loss Adjustment Expense Reserves:

The Casualty Actuarial Society's (CAS) Valuation, Finance, and Investments Committee (VFIC) has prepared the attached note entitled "Materiality and ASOP No. 36: Considerations for the Practicing Actuary".

Actuarial Standard of Practice No. 36, *Statements of Actuarial Opinion Regarding Property/Casualty Loss and Loss Adjustment Expense Reserves*, became effective on October 15, 2000. Among other things, the new ASOP requires the actuary to use the concept of materiality in a number of important ways. The American Academy of Actuary's Committee on Property and Liability Financial Reporting (COPLFR) asked VFIC to prepare a note that would aid the actuary considering materiality in the context of ASOP No. 36.

This note is the result. It is intended to be distributed as an appendix to the Practice Note prepared by COPLFR as well as via the CAS website and *The Actuarial Forum*.

Some of the general concepts of materiality discussed in the note may be relevant beyond statements of actuarial opinion. However, this note does not discuss the intended purposes of analyses in any other contexts, and intended purpose is key to consideration of materiality.

IMPORTANT CAVEAT: This note is intended only as an aid and does not supercede the actuary's professional judgment or the language of ASOP No. 36. Although the note has been prepared by knowledgeable members of VFIC, it has not received the professional review process required for establishment of actuarial standards. Accordingly, the note is not an authoritative document for actuaries and is not binding on any actuary. VFIC recommends that this note be read in conjunction with ASOP No. 36.

2000 Valuation, Finance, and Investments Committee
Casualty Actuarial Society
Harvey A. Sherman, Chairperson

Thomas E. Hettinger	Paul J. Brehm	Gary G. Venter
Donald K. Rainey	Paul B. LeStourgeon	Aaron Halpert
James M. Bartie	Christopher M. Suchar	Evelyn T. (Toni) Mulder
Joseph R. Lebens	Richard W. Gorvett	William M. Wilt
Manalur S. Sandilya	Michael G. McCarter	Kenneth Quintilian

Materiality and ASOP No. 36: Considerations for the Practicing Actuary

Introduction

This note has been prepared by the Valuation, Finance, and Investments Committee (VFIC) of the Casualty Actuarial Society as an aid to the actuary considering the concept of materiality contained in Actuarial Standard of Practice (ASOP) No. 36, *Statements of Actuarial Opinion Regarding Property/ Casualty Loss and Loss Adjustment Expense Reserves.*

ASOP No. 36 requires the actuary to use the concept of materiality in a number of important ways, including:

- determination of whether or not to issue a qualified opinion,
- determination of the need for disclosure of significant risks and uncertainties,
- consideration of factors likely to affect the actuary's reserve analysis, and
- determination of the need for a number of other possible disclosures.

There is no formulaic approach to determining the standard of materiality the actuary should use for a given statement of actuarial opinion (SAO). The ASOP instructs the actuary to evaluate materiality based on professional judgment, any applicable guidelines or standards, and the intended purpose of the SAO.

VFIC intends this note to aid the actuary who must evaluate materiality in the course of preparing a SAO. Following this introduction are three sections:

1. **Materiality and ASOP No. 36:** Discusses the use of the concept of materiality in ASOP No. 36, highlighting its impact on decisions made by the actuary in the course of preparing a SAO.
2. **Materiality in Accounting Contexts:** Reviews the concept of materiality in accounting contexts, including both regulatory and SEC financial reporting. This discussion is not intended to be guidance for the actuary, since an actuary's issues and concerns are not in general the same as those of accountants. Instead, this review is provided to enrich the discussion of potential issues with regard to materiality.
3. **Materiality, Statements of Actuarial Opinion, and ASOP No. 36:** Discusses qualitative and quantitative concepts the actuary may wish to consider while coming to a professional judgment on materiality in the context of ASOP No. 36. Although certain quantitative measures can be suggested for consideration in certain circumstances, no formulaic approach to a quantitative materiality standard can be developed.

Several caveats are in order at this point:

- **This note is intended only as an aid and does not supercede the actuary's professional judgment or the language of ASOP No. 36. Although the note has been prepared by knowledgeable members of VFIC, it has not received the professional review process required for establishment of actuarial standards. Accordingly, the note is not an authoritative document for actuaries and is not binding on any actuary. VFIC recommends that this note be read in conjunction with ASOP No. 36.**

- This note discusses concepts of materiality relevant to the SAO's that are the subject of ASOP No. 36. This note does not focus on considerations of materiality that may be required for other purposes, such as GAAP or Statutory financial statements. Although some of the general concepts of materiality that are discussed here are relevant in other contexts, key to the concept of materiality is consideration of the intended purpose of the analysis. .Discussion of the intended uses of financial statements is beyond the scope of this document.

- ASOP No. 36 applies to any written SAO on loss and loss expense reserves. Many SAO's are prepared to be filed for regulatory purposes with an insurer's statutory annual financial statements. If the actuary is preparing an SAO for some other purpose, e.g., valuation of a company or of a book of business, then the actuary's materiality standards may differ from those relevant to the statutory SAO.

2000 Valuation, Finance, and Investments Committee
Casualty Actuarial Society

Harvey A. Sherman, Chair

Thomas E. Hettinger	Paul J. Brehm	Gary G. Venter
Donald K. Rainey	Paul B. LeStourgeon	Aaron Halpert
James M. Bartie	Christopher M. Suchar	Evelyn T. (Toni) Mulder
Joseph R. Lebens	Richard W. Gorvett	William M. Wilt
Manalur S. Sandilya	Michael G. McCarter	Kenneth Quintilian

Materiality and ASOP No. 36

ASOP No. 36 applies to actuaries issuing written statements of actuarial opinion regarding property/casualty loss and loss adjustment expense reserves in the following situations:

- the opinion is provided to comply with requirements of law or regulation for a statement of actuarial opinion; or
- the opinion is represented by the actuary as a statement of actuarial opinion.

Further, if the actuary's statement includes opinions regarding amounts for items other than loss and loss adjustment expense reserves, ASOP No. 36 applies only to the portion of the statement of actuarial opinion that relates to loss and loss adjustment expense reserves.

Whenever the actuary determines that a material condition exists, the actuary is required to make some response to the condition. The following lists sections of ASOP No. 36 that use the word "material". For convenience, the discussion below quotes some of the context showing how the term ***material*** (with added highlighting) is used in the section.

Again, please note that VFIC has not reproduced ASOP No. 36 in this note. Actuaries should read that document in conjunction with this one.

Sections 3.3.2 d: "The actuary is not required to issue a qualified opinion if the actuary reasonably believes that the item or items in question are not likely to be ***material.***"

Section 3.3.3: "When the actuary reasonably believes that there are significant risks and uncertainties that could result in ***material*** adverse deviation, the actuary should also include an explanatory paragraph in the statement of actuarial opinion." This statement is further clarified. "The actuary is not required to include in the explanatory paragraph general, broad statements about risks and uncertainties due to economic changes, judicial decisions, regulatory actions, political or social forces, etc., nor is the actuary required to include an exhaustive list of all potential sources of risks and uncertainties."

Section 3.4: "… the actuary should consider the purposes and intended uses for which the actuary prepared the statement of actuarial opinion. The actuary should evaluate ***materiality*** based on professional judgment, ***materiality*** guidelines or standards applicable to the statement of actuarial opinion and the actuary's intended purpose for the statement of actuarial opinion."

Section 3.5: "In addition to the reserve methods used, the actuary should consider the relevant past, present, or reasonably foreseeable future conditions that are likely to have a ***material*** effect on the results of the actuary's reserve analysis or on the risk and uncertainties arising from such conditions."

Specific considerations listed in Section 3.5 are the following:

- Coverage Provisions – consider coverage changes, coverage disputes, or coverage litigation.
- Changing Conditions – consider changes in conditions particularly with regard to claims, losses, or exposures that are new or unusual.
- External Conditions – consider forces in the environment that are likely to have a ***material*** effect on the results of the actuary's reserve analysis. However, the actuary is not required to have detailed knowledge of all the economic changes, regulatory changes, judicial decisions, political or social forces, etc., that may affect the settlement values.
- Data – consider whether there are significant data problems or issues.
- Assumptions – consider the sensitivity of the reserve estimates to reasonable, alternative assumptions. When the use of reasonable, alternative assumptions would have a ***material*** effect the actuary should consider the implications regarding the risks and uncertainties associated with such an effect.
- Changes in Assumptions, Procedures or Methods – consider whether the change is likely to have a ***material*** effect on the results. The use of assumptions, procedures or methods for new reserve segments that differ from those used previously is not a change is assumptions, procedures, or methods. Similarly, when the determination of reserves is based on the periodic updating of experience data, factor, or weights, such periodic updating is not a change in assumptions, procedures or methods.

Section 3.7.1 Collectibility: "If the amount of ceded reinsurance reserves is ***material***, the actuary should consider the collectibility of ceded reinsurance."

Section 3.7.4 Risk Transfer Requirements: "... the actuary should ascertain whether an adjustment to the reserves to meet such requirements is likely to have a ***material*** effect on the actuary's reserve analysis or on the risk and uncertainties associated with the reserves."

Section 4.5 Changes in Opining Actuary's Assumptions, Procedures, or Methods: "If a change occurs in the opining actuary's assumptions, procedures, or methods from those previously employed in providing an opinion on the entity's reserves, and if the actuary believes that the change is likely to have a ***material*** effect on the results of the actuary's reserve analysis, then the actuary should disclose the nature of the change. If the actuary can not make a judgement as to whether the change is likely to have a ***material*** effect on the results of the actuary's reserve analysis, the actuary should disclose that there has been a change in actuarial assumptions, procedures, or methods, the effect of which is unknown. No disclosure is required unless the actuary believes that the changes are likely to have a ***material*** effect on the results of the actuary's reserve analysis."

Further, the statement of opinion should include the following disclosure(s):

Section 4.6.a.: "If there have been changes in accounting or processing procedures that significantly affect the consistency of the data used in the reserve analysis and that the actuary believes are likely to have a ***material*** effect on the results of the actuary's reserve analysis, then the actuary should disclose the nature of such changes in accounting or processing procedures."

Section 4.6.c.: "If the scope of the opinion includes consideration of regulatory or accounting requirements regarding risk transfer in reinsurance contracts and if an adjustment to the reserves to satisfy such requirements is likely to have a ***material*** effect on the results of the actuary's reserve analysis, then the actuary should disclose the impact of the risk transfer requirements."

Section 4.6.g.: "If the actuary reasonably believes that there are significant risks and uncertainties that could result in ***material*** adverse deviation, an explanatory paragraph (as described in section 3.3.3) should be included."

Section 4.6.h.: "If the statement of actuarial opinion relies on present values and if the actuary believes that such reliance is likely to have a ***material*** effect on the results of the actuary's reserve analysis, the actuary should disclose that present values were used in forming the opinion … ."

Section 4.6.i.: "If the statement of actuarial opinion relies on risk margins and if the actuary believes that such reliance is likely to have a ***material*** effect on the results of the actuary's reserve analysis, then … ."

Nota bene: The use of ***materially*** in the following excerpt from ASOP No. 36 differs from those discussed above as it refers to the actuary's procedures rather than to the results of the actuary's analysis.

Section 4.8.: The "actuary must be prepared to justify the use of any procedures that depart ***materially*** from those set forth in this standard and must include, in any actuarial communication disclosing the results of the procedures… ."

Materiality in Accounting Contexts

As of this writing, there is no ASOP specifically addressing materiality. Therefore, the primary guidance to the opining actuary is the language in ASOP No. 36. Secondarily, the opining actuary may consider other documents (including this one) originating both inside and outside the actuarial profession.

The NAIC in the preamble to its new Accounting Practices and Procedures Manual (Codification) and the SEC in its Staff Accounting Bulletin (SAB) No. 99 have addressed materiality. These documents discuss materiality from an accounting viewpoint. While neither document can be taken as an Actuarial Standard of Practice, the language itself may provide some understanding as to what constitutes materiality for certain parties interested in the opining actuary's work (e.g., regulators and public auditors).

A. NAIC Accounting Practices and Procedures Manual

The Codification defines a material omission or misstatement of an item in a statutory financial statement as having a magnitude such that it is probable that the judgment of a reasonable person relying upon the statutory financial statement would be changed or influenced by the inclusion or correction of the item.

In narrowing the definition, the following considerations are discussed:

- Some items are more important than others and require closer scrutiny. These include items which may put the insurer in danger of breach of covenant or regulatory requirement (such as an RBC trigger), turn a loss into a profit, reverse a downward earning trend, or represent an unusual event.

- The relative size of the judgment item is usually more important than the absolute size. An example for this is a reserve amount that would significantly impact the earnings of a small company but barely impact the earnings of a large company.

- The amount of the deviation of an item that is considered immaterial may increase if the attainable degree of precision decreases.

B. S.E.C. Staff Accounting Bulletin No. 99

SAB No. 99 uses a similar definition of materiality and has many of the same considerations as does Codification, but it applies to financial statements filed with the SEC.

Of primary importance is that an item that is small in absolute magnitude may be important if its inclusion or modification would change someone's conclusion about the basic financial condition of the company. Numerous examples given in the document include, but are not limited to, masking a change in earnings or other trends, changing a loss into a gain or vice versa, hiding a failure to meet analysts' expectations, and affecting a portion of the business identified as having a key operational role.

But SAB No. 99 notes additional concerns beyond those it has in common with Codification. One issue is that the common practice of using quantitative thresholds as rules of thumb for materiality has no basis in law or accounting literature. Another is that the materiality of items should be considered both separately and in total. An example given considers materiality issues affecting revenues and expenses even though the difference in net income may net out to be small. Similarly, an item may be immaterial in the context of the current year financial statements only to cumulate with other items in the future to yield material differences.

Following are summarized concepts from SAB No. 99 concerning whether a particular set of circumstances is material.

- There should not be exclusive reliance on a percentage or numerical threshold to determine something is material or not.

- The use of a percentage or numerical threshold may provide the basis for a preliminary assumption regarding materiality.

- A matter is material if there is a substantial likelihood that a reasonable person would consider it important.

- Both "quantitative" and "qualitative" factors should be considered in assessing an item's materiality. Experienced human judgment is necessary and appropriate.

Following are qualitative considerations excerpted from SAB No. 99. Note that these items are not necessarily the appropriate items for considering materiality with regard to an SAO submitted to fulfill regulatory requirements. To quote:

> "Among the considerations that may well render material a quantitatively small misstatement of a financial statement item are –
>
> - whether the misstatement arises from an item capable of precise measurement or whether it arises from an estimate and, if so, the degree of imprecision inherent in the estimate
> - whether the misstatement masks a change in earnings or other trends
> - whether the misstatement hides a failure to meet analysts' consensus expectations for the enterprise
> - whether the misstatement changes a loss into income or vice versa
> - whether the misstatement concerns a segment or other portion of the registrant's business that has been identified as playing a significant role in the registrant's operations or profitability
> - whether the misstatement affects the registrant's compliance with regulatory requirements
> - whether the misstatement affects the registrant's compliance with loan covenants or other contractual requirements
> - whether the misstatement has the effect of increasing management's compensation – for example, by satisfying requirements for the award of bonuses or other forms of incentive compensation
> - whether the misstatement involves concealment of an unlawful transaction."

Further, SAB No. 99 concludes that each misstatement should be considered both separately and in the aggregate.

Materiality, Statements of Actuarial Opinion, and ASOP No. 36

VFIC intends that the prior section's review of materiality in an accounting context be regarded as suggestive of issues an actuary may consider in evaluating materiality in the context of ASOP No. 36. One common element between financial reporting and the SAO is that judgments regarding materiality involve both qualitative and quantitative considerations. As noted in Section 3.4 of ASOP No. 36:

> "The actuary should evaluate materiality based on professional judgment, materiality guidelines or standards applicable to the statement of actuarial opinion and the actuary's intended purpose for the statement of actuarial opinion."

Requiring the use of professional judgment and placing importance on intended purpose both emphasize the role of qualitative considerations in evaluating materiality.

Actuaries will naturally also focus on quantitative considerations related to judgments on materiality. No formula can be developed that will substitute for professional judgment by providing a materiality level for each situation. What can be done is to highlight some of the numerical considerations that may be relevant to the determination of materiality in some situations.

A. SAO's Filed with Statutory Annual Statements

Many SAO's are prepared to satisfy the regulatory requirement that such a statement be filed along with a company's annual statement. In that case, a key concern of the management and regulatory audiences for the SAO is company solvency. At least two qualitative issues suggest themselves for consideration in this context:

- Would the item under consideration affect the opining actuary's judgment as to whether the loss and loss expense reserves make a reasonable provision for the liabilities of the entity being opined on?
- Would the item under consideration affect the opinion reader's judgment concerning the impact of the loss and loss expense reserves on the solvency of the entity being opined on, even if the loss and loss expense reserves do make a reasonable provision for the liabilities of the entity being opined upon?

Following are possible quantitative measures that the actuary could consider in the initial phase of determining whether a particular item is material in the context of a SAO prepared for filing with regulators:

- Absolute magnitude of item that represents a correction or a different result if reviewing the work of others.
- Absolute magnitude of item for which data are not available or are incomplete.
- Ratio of item to reserves or statutory surplus.

- Impact of item on IRIS ratios.
- Impact of item on risk-based capital results.
- Likelihood or size of potential variation of ultimate actual result from current expectations.

B. SAO's Prepared for Other Purposes

If the SAO is prepared for a purpose other than that of reporting to regulators, other measures may be appropriate. As a qualitative consideration, the actuary may wish to consider the following issue:

- Would the item under consideration affect the opinion reader's judgment of the impact of loss and loss expense reserves relative to the purpose for which the SAO was obtained?

Here are some other quantitative measures that may be relevant in these contexts:

- Ratio of item to net income or net worth.
- Impact of item on earnings per share.

Evaluation of these quantitative measures to determine a materiality standard must be considered in conjunction with the purpose or intended use of the opinion, the specific circumstances of the entity being opined upon, and the actuary's professional judgment. Variations in a company's circumstances or in the purpose for which the opinion is sought can cause variations in materiality standards even for analyses of otherwise equivalent liabilities.

White Paper on Fair Valuing Property/Casualty Insurance Liabilities

CAS Task Force on Fair Value Liabilities

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Executive Summary

This white paper was undertaken by the CAS Task Force on Fair Value Liabilities in reaction to recent developments by the Financial Accounting Standards Board (FASB) and the International Accounting Standards Committee (IASC). It is meant to be an objective discussion of the issues surrounding the fair valuing of property/casualty insurance liabilities, particularly in the United States. While the recent FASB and IASC proposals are mentioned and quoted, the white paper is meant to be applicable to the "fair value" issue in general, wherever the issue appears.

The paper begins with an introduction and background, including a definition of "fair value." In general, fair value is defined as the market value, if a sufficiently active market exists, or an estimated market value otherwise. Most definitions also include a requirement that the value reflect an "arms length" price between willing parties, so as to eliminate "fire sale" valuations. Most observers agree that a sufficiently active market does **not** exist in most cases for property/casualty insurance liabilities. Hence, estimation methods have to be used to determine their fair value.

A short history of the fair value concept then follows. In brief, the concept of "fair value" gained prominence as a result of the 1980's Savings & Loan crisis in the United States. The accounting rules for these banks at that time did not require the recording of assets at market value, hence, banks were able to manipulate their balance sheets through the selective selling of assets. Troubled banks could sell those assets with market values higher than recorded book values and inflate their reported equity, even as the quality of their balance sheet was deteriorating. The concern was raised that any time financial assets are not held at their economic value (i.e., market or fair value), financial reports can be manipulated through the selective buying and selling of assets.

Since then, the FASB has been embarked on a long-term project to incorporate "fair value" concepts in the accounting for financial assets and liabilities. In December of 1999, they released a document labeled "Reporting Financial Instruments and Certain Related Assets and Liabilities at Fair Value (Preliminary Views)." This document proposed, for the first time, that certain insurance liabilities also be reported at "fair value."

At around the same time, the IASC, in its efforts to develop consistent international accounting standards, released its "Insurance Issues" paper. This paper also proposed a fair value standard for the recording of insurance liabilities.

The paper is organized into the following sections after the introduction

A. **Background** regarding fair value concepts
B. **Fair Value in the insurance context**
C. **Alternatives to Fair Value Accounting** for p/c insurance liabilities.
D. **Methods of Estimating Risk Adjustments** - a brief discussion of possible methods for determining risk adjustments, required in the fair valuing of insurance liabilities. Pros and cons for each method are listed. Detailed discussions of these methods can be found in the technical appendix.
E. **Accounting Presentation Issues**, including alternative income statement or balance

sheet formats in a "fair value" world.

F. **Implementation Issues** surrounding the fair valuing of p/c insurance liabilities for financial accounting statements.
G. **Accounting Concepts**, or how well fair value accounting and the issues discussed in the earlier sections would be viewed in the context of general accounting concepts (such as reliability, relevance and representational faithfulness).
H. **Credit Standing and Fair Value Liabilities**, a discussion of issues related to the reflection of credit standing in determining the fair value of liabilities. This issue has given rise to vigorous discussion, both within and outside the actuarial profession. Due to the controversial nature of this issue, it has been given its own separate section, rather than including it within the earlier sections.
I. **Professional Readiness**
J. **Summary and observations**.
K. **Technical Appendices**.

These sections are meant to be conceptual discussions, with any discussion of detailed implementation procedures left to the technical appendices. The appendices also include a list of references for each section.

Key findings of the task force include:

1. New requirement
In all the accounting conventions that we were aware of, insurance liabilities have not been stated at fair value, resulting in a lack of established practice to draw on. This has implications in numerous areas, including estimation methods, implementation problems and practitioner standards. As with any new requirement, the switch to a fair value valuation standard for property/casualty insurance liabilities would probably result in many unanticipated consequences. These consequences could be mitigated if implementation is phased in. For example, one phase-in alternative would be to institute disclosure requirements at first, followed by full fair value reporting depending on the results of the disclosure period.

2. Alternatives to fair value
There are several alternatives to fair value accounting. These alternatives range from the current use of undiscounted liabilities to conservative discounting approaches to hybrid approaches that combine fair value accounting with other present value methods. Some of these alternatives may result in many of the benefits of fair value accounting, but avoid some of the disadvantages. It is also clear that all approaches have some disadvantages.

3. Expected Value versus best estimate

All the methods discussed in this paper assume that expected value estimates are the starting point in the fair value estimation process. The task force recognizes that confusion sometimes exist as to where current practice stands. While the term "best estimate" is commonly used in current accounting literature, it is not clear whether this means the best estimate of the expected value (***mean***), or the ***mode*** (i.e., most likely value), ***median*** (the value which will be too low half the time, and too high half the time) or ***midpoint*** (the average of the high and low of the range of "reasonable" estimates). While a recent U.S. actuarial standard has cleared up some of this confusion (ASB No. 36, Statement of Actuarial Opinion Regarding Property/Casualty Loss and Loss Adjustment Expense Reserves, discussion of "expected value estimates" and "risk margins"), the task force believes that clarification on this topic within the accounting standards would be beneficial, and would become even more important in a fair value context.

4. Multiple methods

There are multiple methods for estimating the fair value of property/casualty insurance liabilities. All of these methods have their own advantages and disadvantages. No one method works well in all situations. As such, those estimating fair value may need to use a variety of methods. The task force sees a need for any accounting standard to provide for flexibility in estimation methods.

5. Continuum from pricing methods

Several of the possible methods for estimating insurance liability fair values are currently used for pricing. In addition, given that the charged premium may generally be assumed to be a "market" price (in a sufficiently competitive market), that charged premium may be a reasonable initial estimate of the unexpired policy liabilities' fair value. Hence, the initial estimate of a policy's liabilities' fair value may be the result of an existing pricing model.

6. "Typical" line / "typical" company limitation of most current methods

A major issue in determining the fair value of insurance liabilities is the reflection of risk. There are several methods in the current actuarial and financial literature that can be used to calculate this risk margin, for a "typical" line in a "typical" company. Most of these methods will require further development to go beyond the typical line / typical company limitation.

7. A fair value accounting standard would lead to new research

The previous finding discussed a limitation of current fair value estimation methods. The implementation of a fair value accounting standard would lead to new research to address these and other limitations in a fair value estimation process. This would be analogous to the expansion of methods to quantify risk transfer, following the implementation in the United States of FAS 113 (reinsurance accounting).

8. When market prices and "fair value" estimates are in conflict.
The task force observed that there are at least four situations where market prices may be in conflict with the results of a fair value estimation process. In these situations, the fair value estimation process may be preferred over a market value for financial reporting. These situations include:

- *Market disequilibrium.* Given a belief in an efficient market, disequilibrium positions should be only temporary, but how long is temporary? Restrictions on insurance market exit and entry (legal, regulatory and structural) can lead to disequilibrium positions that last years. The underwriting cycle is viewed by some as a sign of temporary disequilibrium, whereby the market price at certain points in the cycle may not equal what some believe to be a fair value.
- *Market disruption.* At various points in time, new events lead to significant uncertainty and temporary disruption in the market for insurance products. Examples can include a threatening hurricane, a newly released wide-ranging court decision and new legislation (e.g., Superfund, or California Proposition 103?). At such times, market prices right after the event may be wildly speculative, or the market may even be suspended, making fair value estimation even more uncertain.
- *Information Asymmetry.* The market price for a liability traded on an active market is likely to be quite different depending on the volume of liabilities actually traded. For example, if a primary insurer cedes 1% of its liabilities, the reinsurers will quite rationally believe that this liability is not a fair cross-section of the primary's entire portfolio: i.e., the ceding insurer is selecting against the reinsurer. Consequently, the price will be rather high, compared to the case where the entire portfolio (or a pro-rata section of it) is transferred. Thus, the "actual market price" is not a better fair value representation than an internal cash flow based measurement unless most of the insurer's liabilities are actually transferred. This situation arises because the market (i.e., reinsurance market) does not have access to the insurer's private information on the liabilities. If all of the private information were public, then the actual market prices for liability transfers would better represent their fair value."
- *Significant intangibles.* Market prices for new business may be set below expected costs for such business, due to the value of expected future renewals. As such, an estimated fair value that ignores this intangible may be materially different from the market price.

Both the IASC and FASB proposals indicate a preference for the use of observed market values over estimated valuations. Given the imbalances noted above, the task force is uncertain as to how to reconcile the realities of the insurance marketplace with the IASC's and FASB's preferences for observed market value. It may be that internal estimates can sometimes be preferable to market based estimates in a fair value accounting scheme.

9. Implications of risk margin approaches without value additivity
Some risk margin methods produce risk adjustments (when expressed as a percentage adjustment) that are independent of the company holding them or the volume of business. Such risk adjustments are said to show "value additivity," i.e., the risk margin for the sum of two items

is the sum of their two risk margins.

Not all risk margin methods result in value additivity. When this is the case, reporting problems can occur. For example, if the risk margin for the sum of line A and line B is less than the sum of the two risk margins, how should this synergy be reported? As an overall adjustment, outside of the line results? Via a pro-rata allocation to the individual lines?

The issue of risk margins and value-additivity centers around discussions of whether markets compensate for diversifiable risk. Diversifiable risk is generally not additive. For example, the relative risk or uncertainty in insuring 2,000 homes across the country is generally less than twice the relative uncertainty from insuring 1,000 homes across the country.

It is not clear whether value-additivity should or should not exist for risk margins in a fair value system. A key question in the debate is the role of transaction costs, i.e., the costs of managing and/or diversifying risk, and how the market recognizes those costs in its quantification of risk margins.

The task force has not taken a final position on this issue. Instead it has flagged the issue wherever it has been a factor in the discussion.

10. Susceptibility to actuarial estimation

We have found nothing in the estimation of fair value that is beyond the abilities of the actuarial profession. We have also found existing models that can be used in the endeavor. This is not to say that the initial results of such actuarial estimation would be problem-free. Problems would undoubtedly occur during any initial implementation, and new techniques and concepts would have to be learned. In short, if fair value accounting rules were implemented for insurance liabilities, actuaries would be capable of producing such fair value estimates, with improvement to be expected over time in both the breadth of estimation methods and actuarial expertise in applying these methods.

11. Increased reliance on subjective assumptions in financial statements

The implementation of fair value accounting for insurance liabilities would increase the number of assumptions underlying reported insurance liabilities. For example, fair value estimates would require assumptions about "market" risk margins and future yields not currently part of the typical property/casualty reserving process. This increased reliance on judgment has been cited by some as a disadvantage of a fair value accounting standard. The task force suspects however that any additional uncertainty caused these additional assumptions is likely to be second order compared to differences in the various company's expected value estimates (before application of risk margins and discounting).

12. Historical comparisons - implementation issues, presentation issues
The implementation of fair value accounting would cause problems with the traditional ways of making historical comparisons, particularly for historic development triangles. One difficulty involves the possible need to restate history, to bring past values to a fair value basis. Should these restated values reflect perfect hindsight, or should some attempt be made to reflect the uncertainty (and estimation risk) that probably existed back then? (Any such restatement may have to consider restating several years of history, based on current reporting requirements.) Or should historic development data not be reported on a fair value basis, similar to current reporting requirements in the U.S. statutory statement, Schedule P, whereby undiscounted values are reported even if the held reserves are discounted?

13. Gross versus net provisions.
Under most accounting systems, both gross and net (of reinsurance) liabilities must be reported. Assuming that the net liabilities contain less risk than the gross liabilities, this would imply the cession of a risk provision. This could change the character of ceded liabilities, as they are currently reported and commonly interpreted.

14. Tax issues.
The change to fair value accounting may have tax implications, where the applicable tax laws rely on financial reporting impacted by the change. Of particular concern is the treatment of risk margins in fair value estimates, relative to tax laws. While risk margins are clearly part of market pricing realities, their acceptance by tax authorities and statutes may not be as clear. This should not be an issue for U.S. property casualty insurers, given the current U.S. tax code, but may have major implications in other jurisdictions.

15. Credit standing reflection in valuing liabilities.
The most contentious issue in the current fair value accounting proposals is whether or not the obligator's credit standing should be reflected in fair valuing its liabilities. Many feel that the existence of guaranty funds, the priority position of policyholders among other creditors in the event of insurer insolvency, and the need for insurers to be seen as solid in order to stay in business make this issue mostly immaterial. There are still strongly held concerns, for those situations where the adjustment may be material. Many feel that the impact of credit standing on liabilities should not be reflected independent on its impact on franchise value, and are concerned that some fair value proposals would fail in this regard. Rather than advocating a certain position, the task force has listed arguments on both sides of this issue.

16. Actuarial workload requirements
Fair value accounting may require reserving actuaries to monitor many more variables than they currently monitor. New items for the reserving actuary to track the impact of may include yield curves, market risk premiums, asset betas, and credit standing. The calculation of the fair value for unexpired in-force policy liabilities may noticeably increase the actuarial workload, relative to the unearned premium and premium deficiency liabilities that they replace. Fair value accounting may also require more frequent "fresh start" updates of estimates than traditional accounting, at least to reflect changing market interest rates.

17. Professional Readiness
Given no established practice in this area to-date, some education effort will probably be required. Professional readiness may also not be determinable until general understanding of the issue increases.

18. Standards versus principles
There is limited amount of practice in this area today. The task force believes that it would be appropriate to first develop general principles or a practice note, and defer development of official standards until practice has had a chance to develop.

The task force hopes this white paper will aid in the understanding of fair value accounting issues as applied to property/casualty insurance. We acknowledge that no one paper can include all that is known about a topic, especially one as new and emerging as this one. As such, we expect this to be only an initial step in the understanding of the issue.

Casualty Actuarial Task Force on Fair Value Liabilities
December 1999 - August 2000

Ralph Blanchard (chair)
Bob Butsic
Catherine Cresswell
Louise Francis
Aaron Halpert
Phil Heckman
Gerry Kirschner
Sarah Krutov
Mike McCarter
Gary Nickerson
Stewart Sawyer
Ernest Wilson
Bryan Young

(other contributors: David Appel, Paul Brehm, Roger Hayne, Gary Josephson, Joe Lebens, Steve Lowe, Glenn Meyers, Elizabeth Smith, Pat Teufel, Gary Venter, American Academy of Actuaries - Committee on Property and Liability Financial Reporting)

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Table of Contents

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Introduction

1) Goal of the paper, and who its authors are.

The following is a discussion of fair value accounting as applied to property/casualty (p/c) insurance liabilities. It is the work product of the Casualty Actuarial Society's Task Force on Fair Value Liabilities, a task force created specifically to address fair value insurance issues raised by several recent accounting proposals (discussed in the Background section below). The issue of possible reporting of insurance liabilities at fair value existed prior to these recent accounting proposals. Hence, this paper is also meant to be a general resource for p/c insurance liability fair value discussions in general.

This paper is not meant to advocate any particular position, but is instead meant to be a "white paper," an objective discussion of the actuarial issues associated with fair value accounting.

2) Scope

The scope of this paper is limited to the issue of fair valuing of p/c insurance liabilities (and related insurance assets), with particular emphasis on insurance accounting in the United States. The analysis includes discussion of estimation issues and their application to accounting. It does not address fair valuing of life or health insurance liabilities, although we recognize the benefits of a consistent approach, where possible, across all insurance liabilities.

The scope is meant to include all material property/casualty insurance liabilities, regardless of the type of entity reporting them in their accounting statements. This would include insurance liabilities held by self-insureds, captives, reinsurers, etc. It would also include unearned premium liabilities, accrued retrospective premium assets/liabilities, material contingent commission liabilities and the like. We have not addressed all possible insurer liabilities, but we have addressed those we believe to be material at an insurance industry level.

3) Format of the paper

The paper is organized into the following sections

A. **Background**, including a definition and history of fair value in general.
B. **Fair Value in the Insurance Context**
C. **Alternatives to Fair Value Accounting** for p/c insurance liabilities.
D. **Methods of Estimating Risk Adjustments** required in the fair valuing of insurance liabilities.
E. **Accounting Presentation Issues**, including alternative income statement or balance sheet formats in a "fair value" world.
F. **Implementation Issues** surrounding the fair valuing of p/c insurance liabilities
G. **Accounting Concepts**, or how well fair value accounting and the issues discussed in the earlier sections would be viewed in the context of general accounting concepts (such as reliability, relevance and representational faithfulness).
H. **Credit Standing and Fair Value Liabilities**, a discussion of issues related to the reflection of credit standing in determining the fair value of liabilities.

I. **Professional Readiness**
J. **Summary and Observations**
K. **Technical Appendices**

These sections are meant to be conceptual discussions, with any discussion of detailed implementation procedures left to the technical appendices.

CAS Fair Value Task Force
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section A - Background

1) Definition of "fair value"

What is "fair value?" Accounting authorities do not currently have a consistent definition for this term. However, a short definition[1] could be:

a. the market value, if a sufficiently active market exists, OR
b. an estimated market value, otherwise.

If no active market exists, an estimated market value can be determined from the market price of similar assets (or liabilities). If no sufficiently similar assets (or liabilities) exist, the estimated market value is based on a present value of future cash flows. These cash flows are to be adjusted for "the effects of ... risk, market imperfections, and similar factors if market-based information is available to estimate those adjustments."[2]

In adjusting these cash flows, one of the more controversial possible adjustments is the impact of the entity's (or obligor's) own credit standing. Under some proposals, the weaker the obligor's financial situation, the lower the fair value of their liabilities would be. The assumption is that the parties to the entity is indebted to would lower their settlement demands, recognizing the risk of possibly getting much less if the entity went insolvent. This would represent a major change to the accounting paradigm for "troubled" companies. A separate section of the white paper has been devoted to this issue, due to its controversial nature and its impact on almost every facet of the fair value discussion.

Note that the fair value is an economic value, but not the only possible "economic value." Other examples of economic values include economic "value-in-use" and forced liquidation value. Economic value-in-use can be defined as the marginal contribution of an item to the overall entity's value. The forced liquidation value is the cash value achievable in a forced sale. Due to the pressures involved, the forced sale price may be materially different from the normal market price.

While fair value accounting could be applied to any asset or liability, it is most commonly an issue for financial assets or liabilities. Financial assets are generally either cash or contractual rights to receive cash or another financial asset.[3] Financial liabilities are generally obligations to

---

[1] There is no universally accepted definition of "fair value" to-date, although they all follow the same general concept given by this short definition. The detailed definition that FASB is proposing can be found in FASB's Preliminary Views document titled "Reporting Financial Instruments and Certain Related Assets and Liabilities at Fair Value," dated December 14, 1999, and labeled "No. 204-B." The definition starts on paragraph 47, with discussion and clarification continuing through paragraph 83. Paragraph 47 states:

> *"Fair value is an estimate of the price an entity would have realized if it had sold an asset or paid if it had been relieved of a liability on the reporting date in an arm's-length exchange motivated by normal business considerations. That is, it is an estimate of an exit price determined by market interactions."*

The IASC has a similar definition (found on page A181 of their Insurance Issues Paper, released November 1999). It reads:

> *"The amount for which an asset could be exchanged, or a liability settled, between knowledgeable, willing parties in an arm's length transaction."*

[2] Paragraph 56 of the FASB Preliminary Views document mentioned above.

[3] This is a simplified definition. A more complete definition includes both options and equities in its scope. Note

provide financial assets.[4]

Lastly, a fair value accounting system focuses on the measurement of assets and liabilities, not income. The income statement in such a paradigm is just a consequence of the changing balance sheet.[5] This is in contrast to a "deferral and matching" approach, such as that used to justify prepaid expense assets (e.g., Deferred Acquisition Costs, or DAC), where the focus is to match revenues and expenses in the income statement. As a result, a fair value income statement could look very different from traditional income statements.

2) Recent history of the fair value concept - United States.

Financial assets and liabilities are accounted for in numerous ways under current U.S. accounting rules (GAAP, statutory insurance and tax). These include historical cost, amortized cost, market value, present value of future cash flows, etc. Each of the various measuring approaches has its advantages and disadvantages. Some approaches produce values that are more readily verifiable than others, but perhaps not as relevant. Others produce more relevant values, if done correctly, but they may not be feasible to use or may be too subject to manipulation.

Historically, many financial assets were accounted for at cost or amortized cost. These values were readily available and verifiable, resulting in balance sheet values that could be produced at minimal cost and that were relatively easy to audit. Likewise, many financial liabilities were at ultimate settlement value, a value that in many cases is contractually set and hence, readily available and auditable.[6]

During the U.S. banking crisis of the late 1980s, this accounting approach caused problems. Banks, which held many financial assets at historical cost, were undergoing financial strains. Many became aware that their reported balance sheet value could be improved by selling those assets with a market value greater than book value, where the book values were based on historical or amortized cost. Assets with market values less than book values were retained, as selling them would only decrease the reported book equity.[7] As a result, many banks were left with asset portfolios dominated by weak and underperforming assets, and many of these banks eventually went insolvent.

---

that this is considered to be a recursive definition, not a circular definition.

[4] This is a simplified definition. A more complete definition would include options-related obligations that would negatively impact the entity if executed.

[5] Accounting systems that focus on the balance sheet are labeled "asset-and-liability-measurement" approaches by the IASC Insurance Issues paper (e.g., paragraph 159). Fair value is an example of, but does not exclusively define, such approaches.

[6] This is clearly not the case for the property/casualty industry, where the amount of the loss is not set by contract, but instead determined via a settlement process.

[7] This process of selling those assets with market-over-book, while retaining those with book-over-market, is referred to as the "cherry-picking" of assets.

The FASB,[8] and many others, felt that a balance sheet based on market values would have provided earlier warning of a bank's financial weakness. They proposed that all bank financial assets be reported at market value, at least for U.S. GAAP financial statements. These concerns resulted in FAS[9] 115, which requires market value accounting for those assets held in a "trading portfolio." These discussions also led to the discussion of fair value accounting for financial assets and liabilities.

New problems arose when determining the scope of FAS 115. Recognizing the fact that many financial institutions compete against one another, whether in the same narrowly defined industry or not, FASB proposed that all U.S. financial institutions be subject to their new asset reporting rules. This would include securities firms, life insurers and p/c insurers (although it is less obvious how p/c insurers compete directly with the others on this list). The FASB's concern was that to not treat all competitors equally in these rules would result in an uneven playing field.

Several parties raised concerns with requiring assets to be held at market value, when the liabilities were not reported at market. They believed that this would cause reported equity to become very volatile and not meaningful. Given the desire for consistency between asset and liability valuation, and the belief by many that market value (or even fair value) accounting for insurance liabilities was not possible, they proposed that the standard's scope exclude the insurance industry. The FASB was not swayed by this argument. They decided to include the insurance industry in the scope of FAS 115, and possibly address the balance sheet inconsistency at a later date.

Since then, the FASB has had a stated vision of having all financial assets and liabilities reported at fair value, pending resolution of any remaining implementation issues.[10]

3) FASB Fair Value project

In 1986, FASB added a broad-based project concerning the appropriate accounting for financial assets and liabilities (i.e., financial instruments) to its agenda. As of a result of the influences mentioned above (and others), it has evolved into the FASB Fair Value project.

The FASB has held discussions on this project during much of 1999. In December of 1999, they issued a "Preliminary Views" document on this project, which was intended to communicate their initial decisions and to "solicit comments on the Board's views about issues involved in reporting financial instruments at fair value." The preliminary views document had a comment

---

[8] Financial Accounting Standards Board, the principal setter of GAAP accounting standards in the U.S. The FASB's standards are superceded only by the Securities and Exchange Commission (SEC). The FASB also must approve AICPA standards of practice before they can become effective.

[9] Financial Accounting Standard. Financial Accounting Standards, or FASs, are issued by the FASB.

[10] In paragraph 3 of the previously mentioned FASB Preliminary View document is a quote from FAS 133, that states as follows. "The Board is committed to work diligently toward resolving, in a timely manner, the conceptual and practical issues related to determining the fair values of financial instruments and portfolios of financial instruments. Techniques for refining the measurement of the fair values of all financial instruments continue to develop at a rapid pace, and the Board believes that all financial instruments should be carried in the statement of financial position at fair value when the conceptual and measurement issues are resolved. [paragraph 334]"

deadline of May 31, 2000.

This FASB document states that insurance obligations settled in cash (which represents nearly all insurance liabilities) are financial instruments, hence, the goal should be to have them reported at fair value. This includes reinsurance obligations. In addition, paragraph 46 of this FASB document "would prohibit capitalization of policy acquisition costs of insurance enterprises." Presumably, the effect of prepaying these expenses would be picked up in the fair valuing of unearned premium liabilities.

As to how to estimate the fair value of these, the preliminary views document references the new FASB Concepts Statement of Present Value-Based Measurements, released February 11, 2000, 2000.

4) IASC - fair value developments and Insurance Issues paper

Concurrent with the FASB developments discussed above, the International Accounting Standards Committee (IASC)[11] has been working to develop standards for financial instruments and for insurance accounting.

Efforts in the area of financial instruments in general include International Accounting Standard (IAS) 39, issued in 1998, and the Joint Working Group on Financial Instruments, currently working to develop a standard by the end of 2000. IAS 39 is very similar to FAS 115, in that it requires investments in a "trading portfolio" to be held at fair value. Unlike, FAS 115, it creates an exception to fair value accounting for any "financial asset ... that does not have a quoted market price in an active market and whose fair value cannot otherwise be reliably measured."[12]

During December 1999, the IASC released an "Issues Paper" focused solely on insurance accounting, with a comment deadline of May 31, 2000.

Among other findings, the IASC paper stated that

- Insurance liabilities should be discounted, and
- If a new international standard is released that requires fair value accounting for financial instruments, then "portfolios of insurance contracts should also be measured at fair value."[13]

(Note that neither the IASC nor the FASB documents, nor their GAAP consequences impact statutory accounting unless the NAIC takes explicit action.)

---

[11] Per the IASC web site as of January 18, 2000 (http://www.iasc.org.uk/frame/cen1.htm), "The International Accounting Standards Committee (IASC) is an independent private-sector body working to achieve uniformity in the accounting principles that are used by businesses and other organisations for financial reporting around the world."

[12] Chapter 30, paragraph 21 of "The IASC-U.S. Comparison Project: A Report on the Similarities and Differences between IASC Standards and U.S. GAAP," Second Edition, published by the FASB in 1999.

[13] These two bullets come from the IASC Issues Paper on Insurance, pages iv-v, bullets (d) and (k).

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section B - Fair Value in the Insurance Context

1) General statement

In general, the fair value projects of both the FASB and IASC propose that any asset or liability that ultimately settles in cash should ideally be valued at its "fair value." This would include (but not be exclusively limited to):

- Loss reserves (i.e., claim liabilities)
- Loss adjustment expense liabilities
- Policy reserves, including unearned premium or unexpired policy liabilities
- Accrued retrospective premium assets
- Return retrospective premium liabilities
- Contingent commissions
- Reinsurance recoverable amounts
- Deductible recoverable amounts
- Salvage and subrogation recoverable amounts.

In addition, a fair value accounting approach (at least according to the FASB) would not recognize prepaid acquisition costs as an asset. Hence, these assets would disappear under fair value accounting.

Premium deficiency reserves would also disappear under fair value accounting, as any expected price inadequacy on in-force policies would be directly reflected in the unearned premium reserve valuation.

Given the absence of an active market for most (maybe all) of these items, their fair value would have to be based on an estimate. The estimate would involve discounted cash flows.

For now, the focus from the FASB and the IASC is on contractual cash flows. Non-contractual cash flows, such as future renewals, would be precluded from the cash flows used to estimate fair value, even when the renewals are largely unavoidable due to existing legal or regulatory rules. The only renewal business flows to be included in these cash flows are those that are contractually guaranteed.[14]

[14] The treatment of renewal business is still an open issue. The quandary these accounting organizations face is that renewal business IS considered currently by those valuing the overall net worth of insurance enterprises. Therefore, a "market value" of the enterprise would include these intangibles. If a market price would include them, then why should a cash flow estimation procedure, generally meant to estimate a hypothetical market value, exclude them? So far, they have leaned against including them, despite a risk of being inconsistent with real-life market valuations, due to problems with reliably estimating the renewal flows.
While both the FASB and IASC proposals include contractually guaranteed renewals in these projected cash flows, the IASC definition further requires that the insurer's pricing flexibility for these renewals be restricted in some fashion.

These discounted cash flows may need to be adjusted for:

- Risk or uncertainty in the flows (with the size of the adjustment based on market compensation for such risk)
- Credit standing of the obligor
- "Market imperfections," including possibly illiquidity.

2) Risk or uncertainty adjustments.

A summary of the October 27, 1999 FASB Board meeting[15] included the following statement, regarding the board's conclusion concerning present value-based measurements.

> "A risk premium is necessary if the risk is identifiable, measurable, and significant. In cases where the risk does not meet those characteristics, risk should not be incorporated into a measurement."

We expect little disagreement that the risk in insurance liabilities is "identifiable" and "significant." We expect the principal discussion to be on the measurability of this risk, in an accounting context.

3) Credit standing of the obligor

As mentioned above, the FASB views the credit standing of the obligor as an integral part of a liability's fair value. After numerous discussions on this topic, they clarified their original statements to say that such credit standing reflection "includes the effect of associated deposit insurance, state guaranty funds, purchased credit insurance, or similar enhancements."[16]

4) Market imperfections, including illiquidity

It is generally recognized that there is no active market for most or all p/c insurance liabilities. Hence, such liabilities will be illiquid to some degree in a fair value context. It is less obvious how a fair value estimate should adjust for such liquidity problems.

5) Alternatives to fair value

Both the FASB and IASC documents recognize outstanding issues regarding the implementation of fair value accounting for insurance liabilities. It is possible they may not be resolved or resolvable in the foreseeable future. Therefore, it is possible that the accounting standards bodies would propose an alternative to fair value accounting, reflecting some of the economics but possible not all that might impact a "fair value."

6) Potential advantages and disadvantages of fair value accounting in the insurance context

Below are some of the advantages and disadvantages to fair value accounting, as it might be applied to insurance liabilities, that have been discussed in prior literature. This partial list is intended to aid in comparing fair value accounting to the various alternatives, discussed in the next section. More detailed discussion of these and other advantages/disadvantages can be found

---

[15] From the FASB Action Alert No 99-35, dated November 3, 1999.

[16] FASB Action Alert, No 99-34, Dated October 27, 1999.

throughout the later sections of this paper.

Potential advantages - Fair Value

- Consistency with assets. If insurance company investments are to be reported at fair value, then its insurance liabilities should be too. This consistent treatment across the entire balance sheet would prevent false volatility in reported earnings and equity.
- Eliminate accounting arbitrage. Valuation of insurance liabilities at other than what they are worth in the market creates incentives to manage earnings through sales of these liabilities, even when done at non-economic prices.
- Consistency with other financial instruments. To the extent that non-insurance financial liabilities are similar to insurance liabilities, they should be accounted for similarly. Otherwise, the inconsistent accounting rules could create competitive advantages based strictly on the accounting, not the economics.
- Relevance. As the value at which such liabilities could be extinguished or traded, fair value should be the most relevant measure for investors.

Potential disadvantages - Fair Value

- Difficulty in measuring. The calculation of reliable fair value adjustments may be a difficult task, and may not always be possible.
- Greater estimation reliance. Fair value accounting systems increase the number of estimates underlying the reported financials. This raises questions as to potential estimation error, and even manipulation of estimates.
- Volatility in earnings. Liabilities held at fair value may show much greater volatility, due to changing yield curves and risk adjustments, versus undiscounted or conservatively discounted liabilities.[17] This additional volatility may provide more noise than information to capital providers and other users of financial statements.
- Cost. Implementation and maintenance of a fair value accounting system will cost time and resources. There may be other alternatives that cost less, and do not have all the disadvantages mentioned above, while still maintaining many of the advantages of fair value accounting.
- Uncertainty. Fair value accounting has never been implemented for insurance liabilities, or other liabilities for which there are no active markets. There will inevitably be some unintended or unexpected consequences from its implementation.

[17] Assuming that the conservative discount rate is not readjusted each reporting period.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section C - Alternatives to Fair Value Accounting

Introduction

For many, the proposals by FASB and IASC present some radically new ways to value balance sheet items and to measure income for insurers. Most of the proposed changes have a reasonable theoretical basis, but a practical implementation of the new methodology will undoubtedly present significant challenges to the actuarial and accounting professions.

For example, as discussed in the *Methods for Estimating Fair Value* section, all of the methods currently available to measure the risk margin suffer from various disadvantages. None of these methods is presently in widespread use for actual valuation of balance sheet liabilities (however, some are commonly used for ratemaking). Although it is likely that more research will evolve given an accounting standard that requires a risk margin, it is difficult to see a route that will arrive at a widely adopted standard approach. Lacking a standard approach (with appropriate guidelines for the magnitude of risk margins by lines of business), it may be difficult to enforce a reliable comparison across insurers.

It is also not clear that all the proposed changes will benefit the industry, its customers or investors. An example is the inclusion of the effect of credit risk in the fair value of liabilities. This requirement implies that an insurer experiencing a lowered credit standing will see its earnings improve. This creates an incentive for companies to increase operational risk and thereby increase the insolvency cost to customers. (For a more detailed discussion of credit risk, see the separate section of this paper on this topic.)

For these reasons, it is prudent to consider some alternatives to the full implementation of the FASB and IASC proposals. The following are alternatives that we have considered or that have been presented in the accounting literature. We do not necessarily endorse any of them, but we list them here in order to enhance the discussion of this topic.

The Alternatives to Fair Value

1. Undiscounted expected value

Use the *undiscounted expected value of the estimated liability payments* as its accounting value. This alternative is essentially the *status quo* for property-liability insurers, although some may have historically used estimates of amounts other than the mean (such as the median or mode). It implicitly assumes that the risk margin equals the discount on the liability. Note that current statutory and GAAP accounting standards allow discounting for some losses (e.g., workers' compensation life pensions). However, the vast majority of liabilities are not explicitly discounted.

The FASB and IASC proposals indicate that the proper way to view the estimation of uncertain cash flows is that the *expected value* of the cash flows is the relevant measurement. Note that the proposals do not directly address this issue with respect to the intended accounting treatment. However, the examples in the documents clearly show the preference for expected value.

The actuarial profession has also recently adopted the expected value criterion. The new Actuarial Standard of Practice No. 36, "Statements of Opinion Regarding Property/Casualty Loss and Loss Adjustment Expense Reserves," specifically requires that the preferred basis for reserve valuation be expected value.

Section 3.6.3 of the ASOP states "In evaluating the reasonableness of reserves, the actuary should consider one or more expected value estimates of the reserves, except when such estimates cannot be made based on available data and reasonable assumptions. Other statistical values such as the mode (most likely value) or the median (50th percentile) may not be appropriate measures for evaluating loss and loss adjustment expense reserves, such as when the expected value estimates can be significantly greater than these other measures." For some, this may be viewed as a change to the previous status quo, while for others, this is merely putting in writing the current practice.

The U.S. regulators' point of view, as expressed in the NAIC Issue Paper No. 55, proposes that the reserves to be booked be "management's best estimate," although the term "best estimate" is not currently defined.

When discussing "expected value" in this paper, we define the term to be without a risk margin, unless stated otherwise.

Advantages

- This is easiest to accomplish. There is no change to current accounting procedures.
- The risk margin equals the amount of the discount, so a risk margin is implicitly included in the liability value.
- The risk margin is directly correlated with the amount of the discount. This is intuitively appealing, since many believe that the amount of risk is positively related to the length of the loss payment tail.
- It is easy to measure the runoff of the liability.

Disadvantages

- It fails to overcome the many problems associated with current accounting, including
  a) Incentive for accounting arbitrage, or transactions undertaken strictly for a favorable accounting result, despite no economic benefit.
  b) Misleading information for decision making, in that transactions that have a poor economic result may look better than those creating a favorable economic result.
  c) Items with significant long-term uncertainty may appear inestimable on an undiscounted basis, even when estimable on a present-value basis.
  d) Companies writing different types of insurance would not be comparable.

- It is a poor calculation of either the risk margin or the present value of the liability. Hence, this alternative results in an accounting value for equity that may not adequately represent the value to investors, policyholders or other parties.

2. Present value at a risk-free interest rate

Use the *present value of the estimated liability payments* as the accounting value. This alternative is equivalent to the fair value, except for the risk margin and adjustment for credit standing. Some would view this as the best practical alternative to fair value, given the difficulties in estimating the risk margin and credit risk adjustment. For some lines of business, such as workers compensation, actuaries routinely calculate present values of the liabilities (although typically using a conservative discount rate). For other lines, the loss and LAE payments patterns needed for present values are usually a by-product of normal loss reserving or ratemaking practices.

This approach is equivalent to *effective-settlement measurement*, discussed on page 22 of the FASB document "Using Cash Flow Information and Present Value in Accounting Measurements" (3-31-99). The effective-settlement method gives the liability value as the amount of assets, which when invested at a specified interest rate, will produce cash flows that match the expected liability cash outflows.

Advantages

- This method is feasible with current actuarial skills and practices. Many insurers currently discount loss reserves for some lines of business. Also, the requisite cash flow patterns are commonly produced in estimating the undiscounted reserves.
- Discounting has widespread acceptance and is fundamental to the life/health industry.
- There is no dispute over how the risk margin should be calculated and applied to individual companies.
- Measuring and displaying the runoff of the liability is not difficult.

Disadvantages

- It will require more work and therefore, expense compared to not discounting.
- A risk margin is not calculated, so the fair value of the liabilities will be underestimated.
- The transition to discounted reserves will expose insurers who have carried inadequate undiscounted reserves that are implicitly discounted (an example is environmental liability). When they are forced to explicitly discount all reserves, some insurers will further discount an already implicitly discounted reserve, rather than admit that the original reserve was inadequate.

- Earnings will emerge closer to the time when the policy is written. (i.e., they are front-ended). This may provide incentives to writing risky long-tail business for companies that have weak earnings.

3. Present value using an industry-standard risk-adjustment

This alternative is similar to #2 above. It uses the present value of expected liability payments as its accounting value, but the present value is taken using a *risk-adjusted* interest rate. Here, *risk-adjusted rate* is defined as a rate that produces a present value *higher* than the present value obtained using the appropriate risk-free interest rate (as in #2 above). To accomplish this, the risk-adjusted rate must be *lower* than the risk-free interest rate. The difference between the two interest rates is called the risk adjustment. For some short-tail liabilities such as catastrophe loss exposure (embedded in unexpired contracts) an adjustment to the interest rate may not be appropriate. In these instances, a *risk margin*, as a percentage of the present value of expected loss, can be added to the present value.

This method is conceptually equivalent to the fair value (with no credit risk adjustment), except that the risk adjustment is determined on an industry-wide basis. Thus, in many cases, the circumstances of the individual insurer would be ignored in favor of accounting simplicity.

There are several approaches that could be applied to determine the industry-standard risk adjustment. A standard-setting organization (such as the AAA or NAIC) could promulgate risk adjustments by line of business or for all lines taken together. The organization might apply some of the methods discussed in Section D and then use judgment to weigh the results in producing the risk adjustment(s). The adjustment could also be set to be the same for all lines, or to vary by line.

Advantages

- It is as nearly as easy as #2 above and it has all of the same advantages plus others.
- It produces a fair value for a typical company's liabilities, since (an) appropriate industry-wide risk margin(s) are (is) provided.
- Comparability between companies may be enhanced, since the risk margins (per unit of like liability) would be the same for each insurer.
- Given the difficulties in accurately estimating risk margins at the industry level in this alternative, it remains questionable whether company-specific fair value estimates would be reliable enough for accounting purposes. Hence, this may be the most practical approach to implementing something akin to fair value.

Disadvantages

- It has the same disadvantages as #2 above except for the omission of a risk margin.
- It may not be a very accurate or reliable calculation of the risk margin for an insurer with atypical liabilities. If risk margins vary by line of business and a single risk margin is

applied to all lines, then insurers writing different types of insurance would not be comparable.
- In the case where line-by-line standards are set, new lines may develop for which no standards yet exist. The standard setters may forever be trying to catch up to market developments.
- There is no formal process to determine the standard-setting body.

4. Mixture of fair value and alternatives

Use *fair value for some liabilities* and one or more of the alternatives for other liabilities. Categories that possibly may require this treatment include unexpired risk (loss embedded in the unearned premium reserve, or UPR), catastrophe losses, environmental losses, ceded losses and loss adjustment expense.

For example, estimating the fair value of UPR runoff can be very difficult when the valuation date occurs as a storm or major catastrophe is threatening, but the public release or reporting of that value is after the event, when the storm either did or did not hit. In this case, an accurate fair value as of the balance sheet date has little relevance at the time losses are reported. Note that retaining the current UPR calculation, and not reflecting fair value until the loss is incurred, would be a "mixture" that retains the current "deferral and matching" paradigm of GAAP accounting.

Under this alternative, either the accounting standard-setting body would establish which categories get which treatment, or the insurer would decide on the basis of a materiality criterion.

Advantages

- This may be the most practical solution to the problems associated with full implementation of the fair value concept.
- This alternative is flexible. It could be amended as actuaries, accountants and other professionals became more adept at measuring the proposed fair value components.

Disadvantages

- It may be difficult to decide which items should get the full fair-value treatment and which items should continue to be valued as they are now.
- It could lead to inconsistent accounting of like items.
- There would be a possibility for accounting arbitrage, or "gaming" the system.
- This alternative could lead to "cliff" changes in liabilities, if a given liability could change valuation standards over its life (such as when the loss component of the UPR becomes incurred).

5. Entity-specific measurement

Use *value-in-use* or *entity-specific measurement.* These measurements substitute the insurer's assumptions for those that the marketplace would make. This measurement would be similar to fair value, but would use an insurer's assumptions regarding interest rate and risk margin. It could also reflect the entity's taxes, servicing cost, affiliate structure and financing costs. Assuming that credit risk were contemplated in the accounting standard, this measure could also incorporate the entity's estimate of the value of its expected default on its obligations. This type of measurement is equivalent to assessing the value at which the entity would be indifferent between running off the liability and settling the liability in a current cash transaction. This value is not necessarily the same as the value that the market would accept for settling the transaction.

In assessing market value of a liability exchange, an important economic effect, called *information asymmetry*, is relevant here. In financial markets, the values of many transactions depend on the amount and quality of information regarding the transaction. Both parties to a market exchange do not always have access to the same information. An example is mortgage lending, where the originator of the loan may have more detailed data on the credit-worthiness of the homeowner than an institution that has purchased the loan. If offered a small portfolio of loans, the loan purchaser will discount the price to guard against anti-selection. However, if the original lender offers its entire portfolio for sale, there is less risk of anti-selection. Therefore, the market value of a single loan chosen at random will depend on how many loans are sold. The same phenomenon will be present for insurance liabilities. In this case, we view the market transaction as an exchange to a reinsurer.

Therefore, in order to satisfy value-additivity in estimating fair value of an insurance liability (where an active market does not exist), one must assume that either the hypothetical market transaction occurs under *symmetric information,* or that the insurer's entire portfolio of liabilities is traded in a market large enough to absorb it. Otherwise, the entity-specific measurement will most likely give a better market value than one obtained by an actual market transaction having a limited size in relation to the entire portfolio.

Advantages

- The insurer would have the most control with this approach.
- An insurer with unique liabilities would be able to use the proper risk margin.
- The method recognizes the current lack of a market for many insurance liabilities, including the large information asymmetry that impedes the existence of an active market. Given this information imbalance, the "market" price is either not transferable to similar liabilities (due to individual portfolio differences), or is a naive price.
- It focuses on the marginal contribution of the item to the total value of the firm, not the exit price for an item for which exit is not a viable alternative. Hence, it may be a more relevant measure to the firm.

Disadvantages

- It might place an additional burden on individual insurers, who would need to derive their specific risk margins.
- It would tend to produce liability values that are not comparable between companies. This would partially defeat the purpose of fair value.
- The method would likely be subject to manipulation by the reporting entity to a greater extent than other alternatives.

6. Cost-accumulation measurement

This approach is discussed on page 22 of the FASB document "Using Cash Flow Information and Present Value in Accounting Measurements" (3-31-99). This method attempts to capture the incremental cost that the insurer anticipates it will incur in satisfying the liability over its expected term. This method typically excludes the markup and risk premium that third parties would incorporate in the price they would charge to assume the liability.

For insurers, these items are the reinsurer's expenses and profit load associated with reinsuring the liabilities. In practice, measurement should be similar to that of the present value alternative (#2) above. Insurers would estimate the liability cash flows and discount them using a prescribed interest rate.

Advantages

- Same as #2.

Disadvantages

- Same as #2.
- It can be dependent on the current corporate structure. For example, it may assume that existing affiliates providing services at marginal cost (to the affiliate) will always be around. This could result in substantial changes in value if the corporate structure changes (e.g., breakup of the parent conglomerate).
- It may not adequately represent what the market would require to transfer the liability.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section D - Methods of Estimating Risk Adjustments

Section introduction and scope

The previous sections discussed general conceptual issues relative to fair value accounting and the principle alternatives to fair value for insurance liabilities. No detail was given as to how the fair value would actually be calculated. This section takes the next step, discussing specific methods that may be used in calculating the fair value of insurance liabilities.

Risk adjustments

Fair value estimates reflect expected cash flows, the time value of money and an adjustment for risk. This section focuses on the last of these components, the risk adjustment. The methods discussed here assume that expected cash flows and risk-free discount rates are already available. For the purpose of all subsequent discussion the starting point for the discount rate before risk adjustment is the risk-free rate.

Risk to the insurer

All the methods discussed here focus on the riskiness of the insured liabilities to the insurer, not the risk that the insurer will default on the liabilities. This latter risk, called credit risk, is very controversial as to its role in estimating the fair value of liabilities. As such, it is being addressed separately, in Section H. Therefore, while some of the methods discussed below may implicitly reflect this credit risk, quantifying that risk is not the intent of this section.

Risk to loss (and loss expense) liabilities

The risk adjustments discussed here generally apply to two major liability categories on the balance sheet: 1) liabilities already incurred (for example, loss reserves) and 2) liabilities not yet incurred for policies already written. The latter liabilities are called the unearned premium (or "unexpired policy") liabilities Although all the other methods we describe for liabilities already incurred could be used for, the unearned premium liabilities, we provide a separate discussion at the end of this section on methods for computing their risk margins.

Other balance sheet insurance items, such as contingent commissions and deductible recoverable amounts may also be subject to a risk adjustment in estimating their fair value. The risk adjustment for these items is not addressed in this section, although some of the methods discussed here may also be feasible for estimating their fair value.

This section begins with a conceptual discussion of risk margins, including a discussion of diversifiable versus nondiversifiable risk. Next, the methods listed below are presented. These presentations are meant to give the reader a brief conceptual overview of the methods (a more involved discussion is included in the appendices). At the end of this section, a chart comparing the listed methods is provided.

(Note: Neither the inclusion of a method, exclusion of a method, nor the order of the methods listed is meant to imply any preference or priority by the task force. Methods were listed if members of the task force felt it deserved consideration, whether or not consensus was achieved.)

1) Capital Asset Pricing Model (CAPM) based methods, where the liability beta is calculated from insurers' asset and equity betas.

2) Internal Rate of Return (IRR) method, where the risk adjustment is derived from cash flow and rate of return on equity (ROE) estimates.

3) Single Period Risk-Adjusted Discount method, where the calendar year ROE is used to find a risk adjusted interest rate.

4) Methods that use historical underwriting results to derive a risk adjustment.

5) Methods using probability distributions of aggregate losses.

6) Determining fair value estimates from reinsurance transactions.

7) Direct estimation of liability market values based on share prices of property-liability insurance companies.

8) Transformed distribution methods, where the probability distribution of liability outcomes is altered to produce a higher expected value.

9) Naive methods using rules of thumb.

10) Other methods.

Conceptual overview - risk margins

The IASC (paragraph 243) and FASB (Concept statement 5 paragraphs 62 - 71) documents require the use of a risk margin when measuring the fair value of an uncertain liabilities (such as an insurer's liabilities) by discounting the expected liability cash flows. The finance and actuarial literature generally support this approach. (Butsic, Cummins, D'Arcy, and Myers-Cohn.)

The economic rationale for a risk margin is that a third party would not accept compensation for a transfer of liabilities if such payment reflected only the present value of the cash flows at a risk-free interest rate. The acquiring entity would get an expected risk-free return while bearing risk. A market exchange of the liability would therefore require a premium or risk margin over and above the present value of the liability discounted at the risk-free rate.

In this section we discuss various possible feasible methods for estimating a risk margin. All of these methods have been used for estimating risk margins, either for direct application to balance sheet liabilities or in ratemaking. Financial theory indicates that the same principles for estimating the risk margin in pricing would also apply to a fair valuation of outstanding liabilities. For certain kinds of short tail liabilities, such as claim liabilities associated with catastrophes, the risk margins for pricing may be much larger than the risk margins for liabilities, however. This is because, once a catastrophe has occurred the uncertainty regarding future payments may be relatively modest, compared to the quite large level of uncertainty before the event has occurred.

There are two major paradigms used to compute risk loads that are represented in this paper: the finance perspective and the actuarial perspective. These two paradigms differ in their treatment of diversifiable versus nondiversifiable risk. In the context of liability fair value, diversifiable risk is defined as risk that can be reduced, per unit of liability volume, as more volume is added. For example, if two statistically independent risks are combined, their joint risk will be reduced due to the tendency of bad outcomes from one being offset by good outcomes in the other. In contrast, nondiversifiable (or systematic) risk is defined as risk that cannot be reduced, per unit of liability volume, as more volume is added. Here, bad or good outcomes in one risk are matched with the same result in the other.

The amount of diversification depends on the correlation between the units being added. This

$$\sigma(x+y)=\sqrt{(\sigma_x^2+\sigma_y^2+2\rho\sigma_x\sigma_y}$$

effect is evident in the square root rule for summing standard deviations:
Where ρ is the correlation between x and y, σ_x is the standard deviation of x and σ_y is the standard deviation of y.

Adding more units to a portfolio may or may not reduce its risk. If the correlation between the units is one, then there is no reduction in risk per unit volume from adding more of the units. In this case the standard deviation of the sum will equal the sum of the standard deviations, and when this is normalized by dividing by the mean of the portfolio, the risk per unit is unchanged. In investing, for instance, adding more shares of a given company's stock to one's portfolio will not reduce the portfolio's risk, since the shares added will be perfectly correlated with the shares the investor already owns.

If the correlation between the units is less than one, then there is a reduction in risk per unit volume from adding the units. Thus, if an investor adds to the portfolio shares of a company not already in it, the risk should decline since the correlation of the new stock with stocks in the portfolio should be less than one. If the correlation is negative then there can be a significant reduction in risk.

An example of diversifiable risk from insurance is the random occurrence of losses — where the fortuitous amount of one claim does not influence the amount of another claim. An example of nondiversifiable risk from insurance is medical inflation, where a change in the cost of medical care will simultaneously effect the value of general liability and workers compensation reserves. Another example is parameter risk, where the mean (or other parameter) of a loss distribution is unknown. Here the uncertainty in the mean affects all losses included in the distribution.

The finance perspective:

The classical finance perspective, as reflected in such methods as CAPM and internal rate of return, posits that an investor is compensated only for that risk that is not diversifiable. Diversifiable risk is not rewarded in the financial markets, because an investor can eliminate this risk by holding a diversified portfolio of securities. The finance perspective quantifies nondiversifiable risk, which is also called systematic risk, by measuring the correlation of a security's return with the market's return. From the finance perspective, if an investor owns a sufficiently diversified portfolio of securities, the only portion of the securities' return that cannot be diversified away is due to its co-movement with the market. Thus, much of the finance literature tends to treat systematic risk and covariance with the stock market as synonymous, and ignores other possible approaches to defining and quantifying diversifiable risk.[18] For determining risk loads in insurance, this may translate into measuring the correlation between insurance companies' returns from underwriting and market returns on its shareholder's equity.

The actuarial perspective:

In determining risk loads, what has come to be known as the actuarial perspective, in general, looks at the contribution of a policy to the total risk of the enterprise. (Risk loads based on aggregate probability distribution reflect the actuarial perspective.) The contribution to total risk will have a component that is diversifiable (process risk) and a component that is nondiversifiable (parameter risk). For many lines, especially in large insurers, the component due to process risk will be small, however, due to the law of large numbers. The actuarial perspective views the nondiversifiable or parameter risk component as that portion of total uncertainty due to the enterprise's inability to accurately measure its true liability and expense costs. While parameter risk may sound analogous to systematic risk, as both are viewed by their users are nondiversifiable, they are different concepts. Systematic risk is measured by calculating correlations with market returns. Parameter risk, where quantified, is measured through the use of Bayesian statistics.

---

[18] Certain approaches, such as Arbitrage Pricing Theory allow factors other than beta to be used in the quantification of risk. Except for some very recent research work, these approaches have not influenced the finance-based methods used to compute risk loads in property and casualty insurance.

The characterization of the finance approach as quantifying only nondiversifiable risk and the actuarial approach as including both diversifiable and nondiversifiable risk is an oversimplification. Stulz[19] points out that in the real world, total risk often matters, and costs incurred by companies to control total risk are rewarded in the financial markets and the failure to do so may be punished. For some kinds of insurance, such as catastrophe insurance, it could be difficult to find a market unless some kinds of "diversifiable" risk were rewarded. Property catastrophe risk is diversifiable in a perfect market, but the mechanisms for doing so are so costly that in practice it is only partially diversifiable. As in the case of formally nondiversifiable risk, the whole industry is in the same boat, so the market treats the risk as systematic and policyholders in catastrophe-exposed areas pay a risk premium for insurance coverage. If an efficient means of diversification were to arise, then that situation would change.

While the actuarial based methods often explicitly incorporate process (diversifiable) and parameter (nondiversifiable) risk components into the risk load formulas, some of the finance-based methods, such as internal rate of return, may implicitly incorporate this risk as part of the total return on equity required by an insurance company.

The discussion surrounding diversifiable versus nondiversifiable risk is still evolving. The reader should be aware that differing views exist as to whether only diversifiable, or both diversifiable and nondiversifiable risk should be included in risk adjustments. The reader should also be aware that there are also very different approaches to measuring the nondiversifiable component.

[19] Stulz, Rene, "Whats wrong with modern capital budgeting?", Address to the Eastern Finance Association, April, 1999

Method 1. The CAPM Approach

(Note: references to specific authors mentioned below and in the discussion of subsequent methods can be found in the Appendix.)

CAPM is the method used in Massachusetts rate filings in the Automobile and Workers Compensation lines. Myers and Cohn developed the underlying theory.

The method equates the present value of the premium to be charged on a policy to the present value of the losses plus the present value of the underwriting profits tax plus the present value of the tax on invested surplus and premium.

$$PV(P) = PV(L) + PV(UWPT) + PV(IT),$$

where P = Premium, net of underwriting expenses
L = losses plus loss adjustment expenses
$UWPT$ = underwriting profits tax
IT = tax on investments

Losses are discounted at a risk-adjusted rate. The premium portion of underwriting profits is discounted at a risk-free rate and the liability portion is discounted at a risk-adjusted rate. Investment tax is discounted at the risk-free rate. The risk-adjusted rate used in the calculations is derived from CAPM.

$$r_L = r_f + \beta_L(r_m - r_f)$$

where r_L = risk-adjusted rate
r_f = one period risk-free rate
$\beta_L = \mathrm{Cov}(r_L, r_m)/\mathrm{Var}(r_m)$ = the liability or underwriting beta
r_m = expected rate of return on market portfolio

β_L, the underwriting beta, is a measure of the covariance between the underwriting profits for a line of business and the stock market. It represents the systematic risk to the insurer for writing the policy. Note that β_L is usually considered to be negative. Otherwise insurance companies would incur exposure to risk for a reward equal to or less than the risk-free rate, an illogical conclusion.

Although the Myers-Cohn approach is typically applied in ratemaking to compute risk adjusted premiums for new policies, the risk-adjusted discount rate from the calculation can be used to discount outstanding reserve liabilities as well.

There are at least three approaches to computing β_L. The first method is broadly similar to the direct estimation technique (Method 7 of this section). Here, a time series of publicly traded insurer data is analyzed. A beta of equities is determined from insurance company stock prices. A beta of assets is determined from a weighted average of insurance company asset betas. The liability beta is determined by subtracting the asset and equity betas, weighted by their respective leverage values. The risk margin, as a reduction of the risk-free rate, equals the liability beta times the market risk premium. This is the method used in Massachusetts.

The second method uses accounting data to measure the covariance between insurance underwriting returns and the market.[20] A third CAPM-based approach measures beta for a line of business by quantifying the covariance of that line's underwriting return with the return for all property and casualty lines.[21]

A numerical illustration of the method is shown in the Appendix.

Advantages

- The method has actually been done. In Massachusetts it is the standard method used in the workers compensation and personal auto, with risk margins being positive and stable. Note that this has only been applied to lines that are relatively homogeneous, and where public data is generally available.
- The method is objective and the analysis is reproducible.
- The method has been in use for over a decade and has been reviewed by many economists.

Disadvantages

- Several stages of estimation can produce measurement errors.
  a) Some insurers in the data are also life insurers; carving them out requires estimating the equity beta of the life operation.
  b) The liabilities may be under- or overstated in the financial statements.
  c) Mutual insurers, nonpublic companies, self insurers and captives are not included in the analysis, introducing a potential bias
- Intangible assets like franchise value could distort the results. Another similar problem is that the present value of income taxes is embedded in the liability value and cannot be easily separated from it.
- Measurement errors on the beta for assets have a leveraged effect on the measurement of underwriting betas.
- It relies on the CAPM model, which may not accurately predict returns for insurance firms, as discussed below.

---

[20] Kozik, Thomas, "Underwriting Betas-The Shadows of Ghosts," Proceedings of the Casualty Actuarial Society (PCAS) LXXXI, 1994, pp. 303-329.

[21] Feldblum, Shalom, "Risk Load for Insurers", PCAS LXXVII, 1990, pp. 160- 195

The CAPM beta has come under considerable criticism recently in the finance literature. CAPM only recognizes nondiversifiable risk, assuming an efficient, friction-free market. The magnitude of transaction costs to diversify an insurance portfolio violates the friction-free assumption, casting doubt as to the applicability of CAPM to valuing insurance liabilities.

Fama and French have shown that factors other than beta contribute significantly to the explanation of company stock returns.[22] Their work has caused a great deal of discussion in the finance community about the use of CAPM and beta for estimating equity returns and computing cost of capital. Alternatives to CAPM that look CAPM-like but incorporate factors other than beta into the determination of the risk premium have attempted to address some of the deficiencies of the CAPM model. For instance, Fama and French have presented a method for deriving costs of equity that uses two additional factors as well as beta.[23] Some of the models that appear to be generalizations of CAPM and use factors other than beta are better known as examples of the Arbitrage Pricing Model. An introduction to this more general approach is provided by D'Arcy and Doherty.[24]

Members of the actuarial community (as opposed to members of the finance community) have also criticized CAPM approaches. Much of the criticism focuses on the unreliability of estimates of underwriting betas as opposed to estimates of equity betas examined by Fama and French. Kozik[25] notes that a number of authors have measured the underwriting beta to be zero or negative (i.e., no risk load necessary on insurance). He provides a detailed discussion of the flaws in current methods of measurements of the underwriting beta, which can cause such results to be obtained.

Note that much of the underlying theory of CAPM is widely used and accepted, although the actual mechanisms for measurement have been criticized. Some of the criticisms of CAPM have been addressed in extensions of CAPM such as contained in the Automobile Insurance Bureau's Massachusetts Rate Filing (1998). Extending CAPM to address some of its limitations is currently an area of active research.

It should be noted that many of the limitations of the CAPM approach may apply to other methods presented in this paper, whenever those methods use CAPM to determine a rate of return.

---

[22] Fama, Eugene and French, Kenneth, "The Cross Section of Expected Stock Returns" Journal of Finance, Vol 47, 1992, pp. 427-465

[23] Fama, Eugene and French, Kenneth, "The Cross Section of Expected Stock Returns" Journal of Finance, Vol 47, 1992, pp. 427-465

[24] D'Arcy, S. P., and Doherty, N. A., "The Financial Theory of Pricing Property-Liability Insurance Contracts," Huebner Foundation, 1988

[25] Kozik, Thomas, "Underwriting Betas-The Shadows of Ghosts," Proceedings of the Casualty Actuarial Society (PCAS) LXXXI, 1994, pp. 303-329.

The Pricing-Based Methods (Methods 2 and 3)

Under this general category of methods, the fair premium for a group of polices (which could be those of a line of business or an entire company) is first determined. In this calculation, the value of all nonliability premium components (such as commissions and general expenses) is excluded from the fair premium calculation. The resulting premium amount, by definition, is the fair value of the liability (losses and loss adjustment expenses). Since the liability fair value and its expected payments are known, the implicit risk-adjusted interest rate at which the payments are discounted can be readily found. Subtracting this value from the risk-free rate gives an estimate of the risk adjustment to the risk-free rate. Note that this approach can be used to compute a dollar-value risk load (to apply to liabilities discounted at the risk-free rate) rather than an adjustment to the discount rate.

This method can be applied to any prospective pricing model that uses expected cash flows. The most prevalent cash flow approaches are the internal rate of return (IRR) and the risk-adjusted discount (RAD) models.

It should be noted that the standard pricing-based methods give a risk margin that is a composite of the risk characteristics of liabilities already incurred and the unexpired policy liability. As the time since policy issuance increases, there may be a significant information gain in a book of liabilities (e.g., the insurer knows more about claims once they are reported) This effect is most pronounced for property insurance with significant catastrophe potential. To separately measure the risk margins in the reserve and unexpired policy portions of the insurer's liabilities, the pricing methods can be modified. For example, in the IRR model, the capital requirement and/or the required ROE may be different per unit of liability for the two liability types.

Method 2 - The IRR method

The IRR method is used by the NCCI in workers compensation rate filings.[26] It does not directly produce a risk margin, but it can easily be adapted to do so. The underlying theory is standard capital budgeting.[27]

Under the IRR method, a cohort of policies, written at the same time, is modeled over time until all claim payments are made. At each stage (usually quarterly or annually) the cash flows (premiums, losses, expenses, income taxes and investment returns) and balance sheet values are estimated. Capital is added based on capital allocation rules, frequently as a fixed proportion to liabilities. The application of these capital allocation rules results in an initial amount of capital,

---

[26] Cummins, J. David, "Multi-Period Discounted Cash Flow Ratemaking Models in Property-Liability Insurance," Journal of Risk and Insurance, March 1990, Volume 57:1, pp. 79-109.

[27] Brealy, Richard A. and Stuart C. Myers, 1996, "Principles of Corporate Finance (5th Edition)", McGraw-Hill, New York

then a subsequent capital flow, based on the amount of additional or withdrawal of capital necessary to maintain the capital allocation assumption at each point in the policy flows.

When the internal rate of return on the capital contributions and withdrawals equals the required rate of return on the capital (equity), then the fair premium is obtained.

The inputs to the IRR method are the capital allocation rules (e.g., the required amount of equity per unit of liability), the expected payments pattern of the policy flows, the investment return on cash flows, the income tax rate and the required return on equity. Note that the expenses and the premium cash flows need not be included in this calculation, since we are only trying to value the liability itself.

The required ROE can be determined using a variety of approaches. A simple approach often used by insurance companies is to select a rate of return based on examining actual historical rates of return on equity for insurance companies. Roth advocates this approach.[28] Another approach is to use CAPM to estimate the industry-average insurer equity beta and then to derive the appropriate ROE, given beta. An alternative way to estimate the required ROE is to use the dividend growth model, which has been documented in rate filings. Still another approach might use the "hurdle rate" for an insurer that is derived from its experience raising capital.

The required capital could be based on the company's internal capital allocation rules. Absent this, industry-wide "rules of thumb" or rating agency dictated norms might be used. Note that the capital typically used in this calculation is "required" or "target" capital, not actual capital. Care must be taken where the capital allocation assumption is dependent on the required ROE assumption.

An additional complication arises where fair value rules require the use of "market assumptions" wherever possible, over individual company assumptions. This could imply that the capital allocation rules that drive the market price (if one can be said to exist) should be used instead of the company's own internal capital assumptions.

The investment return under a fair value paradigm typically is the set of currently available market yields for investments. This may be complicated by investment in tax-exempt investments, especially where the company has significant tax advantages or disadvantages relative to the market. Many users of IRR models make the simplifying assumption that all investments are made in taxable securities.

A numerical illustration of the method is shown in the Appendix.

---

[28] Roth, R., "Analysis of Surplus and Rates of Return Using Leverage Ratios,", 1992 Casualty Actuarial Society Discussion Paper Program - Insurer Financial Solvency, Volume 1, pp 439-464

Advantages

- The IRR is commonly used to price insurance products. The extension to calculate risk margins is straightforward and will produce positive and stable risk margins.
- The method is conceptually simple and easy to explain.
- The method is objective and the analysis is reproducible.
- The method will work at the individual insurer level.

Disadvantages

- All of the methods for determining the required return on equity have problems and they can produce different answers:
  a) A required ROE based on historical returns depends on the historical period chosen.
  b) A required ROE based on CAPM is subject to the limitations and criticisms that apply to CAPM (see Method # 1 above).
  c) The dividend growth method requires some subjective estimation — it will not work for companies with erratic or no dividends.
  d) Internal management "hurdle" rates, based on a company's experience in raising capital, are very subjective and may not be consistent with the market value approach under fair value.
- The number of steps required makes this a fairly indirect method.
- Estimating the present value of income taxes requires a modification to the method.
- A required capital estimate is needed. There is no agreed upon method for doing this, and no consensus as to whether it should be the company's or the industry's capital allocation or requirement.

Method 3 - The Single-Period RAD (Risk-Adjusted Discount) method

This method shares some features of the above IRR method. It is based on the risk-adjusted discount method.[29,30] Here the relationship between the required ROE, the expected investment return, the income tax rate and the capital ratio is used to find the implied risk-adjusted interest rate. Like the above IRR method, the balance sheet values are fair value quantities. It is simpler than the IRR model since the risk adjustment is derived directly from a formula (shown in the Appendix), rather than by an iterative process.

The inputs to the single-period RAD method are the required amount of equity, the investment return on cash flows, the risk-free rate, the effective income tax rate and the required return on

[29] Butsic, Robert, "Determining the Proper Discount Rate for Loss Reserve Discounting: An Economic Approach," 1988 Casualty Actuarial Society Discussion Paper Program - Evaluating Insurance Company Liabilities, pp. 147-188.

[30] D'Arcy, Stephen P., 1988, "Use of the CAPM to Discount Property-Liability Loss Reserves", Journal of Risk and Insurance, September 1988, Volume 55:3, pp. 481-490.

equity. The required ROE can be determined using one of the methods described above for the IRR approach. The required capital and the investment return are estimated using historical industry data, or from one of the alternative methods described above for the IRR approach. Note that the required capital needs to be consistent with the fair value of the liabilities. For example, if the fair value of reserves were less than a non-fair value such as ultimate undiscounted liabilities, the required capital would go up.

The simplicity of this method arises from the assumption that the risk adjustment (as a reduction to the risk-free rate) is uniform over time. Thus, evaluating an insurance contract over a single period will be sufficient to determine the risk adjustment. To illustrate the method, we assume the following:

- capital is 50% of liability fair value,
- required ROE is 13%,
- expected investment return (EIR) is 7%,
- risk-free rate (RFR) is 6%,
- income tax rate is zero, and
- fair value for the liability is $100 at time zero.

The formula for the risk adjustment is:

risk adjustment = capital ratio x (ROE – EIR) + RFR – EIR

= 0.02 = 0.5 x (0.13 – 0.07) + 0.06 – 0.07

The formula for the resulting risk-adjusted interest rate is:

risk-adjusted interest rate = RFR - risk adjustment

= 0.04 = 0.06 - 0.02

To see that this works, note that the beginning assets are the fair premium for the liability of $100 plus the required capital of $50. This amount grows to $160.50 (i.e., $150 x 1.07) at the end of the year. The expected amount of liability grows at the risk-adjusted rate of 4% to $104. Subtracting this amount from assets gives $56.50, which represents the required 13% return (56.5 / 50 = 1.13) .

The income tax rate, however, is not zero, so the formula for the risk adjustment (see the Appendix) is somewhat more complicated than shown here. The Appendix provides the complete formula and also gives a numerical illustration of the method.

Advantages

- The method is very simple and transparent. It is easy to explain and to demonstrate with a spreadsheet.
- The method is reliable, robust and will produce positive and stable risk margins.
- Inputs are presently available from published sources. For example, many rate filings with state insurance departments have estimates for required ROE and capital leverage.

Disadvantages

- The method will only produce an industry-average or company-average risk adjustment (to the risk-free rate). It would be difficult to apply the method to produce specific lines of business risk adjustments.
- This method has the same disadvantages relative to the selected ROE as the IRR method.
- This method has the same disadvantages relative to the selected "required capital" as the IRR method.

Method 4 - Methods Based on Underwriting Data

A pragmatic approach to developing liability risk adjustments is to use published underwriting data. Over a sufficiently long period of time companies are assumed to earn enough in profit on the policies they write to be adequately compensated for the risk they bear. This method assumes that the historical returns indicate the true market perception of the fair profit for bearing insurance risk. The historic profit or risk load can then be related to the risk adjustment required for discounted liabilities.

Typically, risk adjustments based on underwriting data use information published in insurance companies' annual statements. To obtain stable results by line of business applicable to a typical company, data aggregated to industry level by sources such as A. M. Best can be used.

The published literature on risk adjustments using underwriting data primarily focuses on estimating a risk adjustment to the factor used to discount liabilities. Alternative methods for computing risk-adjusted discount rates use a CAPM approach to compute the risk adjustment.

Although we focus on using underwriting data to compute risk-adjusted discount rates, the same data can be used to derive an additive risk load instead.[31] Risk adjustments incorporated through the discount rate are discussed first, followed by discussion of risk adjustment via an additive risk load.

Using Underwriting Data to Adjust the Discount Rate

Butsic introduced the concept of using risk adjusted discount rates to discount insurance liabilities.[32] He argued that a liability whose value is certain should be discounted at a risk free rate. The appropriate risk free rate to use for the certain liabilities is the spot rate for maturities equal to the duration of the liabilities. If certain liabilities are discounted at the risk free rate, then uncertain liabilities should be discounted at a rate below the risk free rate. The formula for the risk-adjusted rate is:

---

[31] There are several different ways to make a risk adjustment. One way is through an additive risk load to the otherwise calculated present value estimate (based on risk-free discount rates). A second is by discounting the expected cash flows using a risk-adjusted discount rate. A third is by adjusting the individual expected cash flow amounts for each time period, replacing each uncertain amount with the certainty equivalent amount (i.e. the fixed amount for which the market would be indifferent between it and the uncertain amount being estimated.) A fourth is by adjusting the timing of the estimated cash flows (sometimes used when timing risk is thought to dominate amount risk).

[32] Butsic, Robert, "Determining the Proper Discount Rate for Loss Reserve Discounting: An Economic Approach," 1988 Casualty Actuarial Society Discussion Paper Program - Evaluating Insurance Company Liabilities, pp. 147-188.

$$i_L = i - e\,(R - i),$$

where i_L = the risk-adjusted discount rate for liabilities,
i = the risk free rate for duration equal to the duration of the liabilities,
e = a leverage factor, equal to surplus divided by the present value of liabilities,
$(R - i)$ = the market risk premium, i.e., the excess of the market's return over the risk-free rate. The market return is usually measured as the return on a stock market index such as the S&P 500 or the return for all NYSE stocks, but other interpretations are possible.

The above term "$e\,(R - i)$" represents the adjustment to the risk free rate for the riskiness of the liabilities.

There is an analogy between this formula and that for a company's cost of equity based on the CAPM.

$$i_E = i + \beta_c\,(R - i)$$

where i_E = the cost of equity for a company,
i = the risk-free rate,
β_c = the company's beta, based on the covariance between the return on the company's stock and the market's return,

The specific procedure for computing the adjustment is described in detail in the Appendix.

Note that the method's results can be very sensitive to the historical time period used as the source of the underwriting data. For example, the selection of an historical period that includes a major market disruption, such as a workers' compensation crisis, major catastrophe, or mass tort eruption, can produce drastically different indications than a time period that excluded this major disruption. Thus, it is necessary to consider how long a time period is required to obtain stable and reasonable results and whether the method is unstable over time. The longer the historical period used for computing the risk adjustment, the more stable the results will be, but the less likely they are to reflect current trends in the underwriting cycle or business environment. The shorter the historical period used, the more likely it is that the adjustment will reflect the current environment, but at a cost of being more unstable and more susceptible to infrequent random events such as catastrophes (or the short-term absence of the long-term catastrophe or large loss risk).

An additional effect that must be considered is the effect of taxes. As shown by Myers and Cohn[33]

[33] Myers, S and Cohn, R, "A Discounted Cash Flow Approach to Property-Liability Rate Regulation," Fair Rate of Return in Property-Liability Insurance, Cummins, J.D., Harrington S.A., Eds, Kluwer-Nijhoff Publishing, 1987, pp. 55-78

and Butsic[34], taxes increase the premium needed to obtain a target rate of return and therefore decrease the effective risk-adjusted discount rate. This effect is embedded in the data used to derive the risk-adjusted discount rate. It might be desirable to segregate this effect from the pure risk adjustment. A procedure for doing this is discussed in the Appendix.

Advantages

- The approach produces an adjustment to the discount rate without requiring the computation of a liability beta. As discussed above in the CAPM method for estimating a risk adjustment, the liability beta is one of the more controversial features of the CAPM approach.
- The approach does not require the computation of a leverage ratio
- The approach is relatively easy to implement. Spreadsheets can be placed on a web site containing a sample calculation
- The data required, such as Bests Aggregates and Averages, is relatively inexpensive and readily available
- A paper presenting the approach has been included in the syllabus of the Casualty Actuarial Society for over 10 years. A description of this technique is, therefore, readily accessible to actuaries (or anyone else who accesses the CAS web site.)
- This method can easily be applied to individual lines where annual statement data is available.

Disadvantages

- Results can be very different depending on the historical time period used. This committee's research indicates that changing the time period used for the calculation in one instance changed the all-lines risk adjustment from 4.5% to 1.0%. The committee believes that the results for recent historical periods reflect certain well-known market disruptions such as the impact of the recognition of asbestos and environmental liabilities. Also, the industry has been in a protracted soft market, which has depressed underwriting profitability in the recent historical data.
- Results for a single line can be unstable. Some lines are unprofitable for extended periods of time and this method may not produce a positive risk load. Useful data for lines with very long tails (or without industry data available) may be a problem. Examples of such include medical malpractice-occurrence and directors & officers (D&O, for which industry accident year data may not be available).
- Pricing adequacy may vary by line based upon individual line characteristics such as regulatory environment, market conditions, geography, etc. An impact of this is cross subsidization of lines where some lines are undercharges at the expenses of other lines. Thus the results for a single line, even over relatively long time periods can be misleading. (Our research showed that at least one regulated line had a negative risk adjustment using this approach for 30 years.)

---

[34] Butsic, Robert P., 2000, Treatment of Income Taxes in Present Value Models of Property-Liability Insurance, Unpublished Working Paper.

- Results will be affected by "smoothing" in published financial numbers.
- The method requires accident year data to do the computation correctly, or else it is susceptible to distortion from events with long-term latency issues, such as mass torts or construction defect.
- Results using individual company data may be too volatile, hence, the method has usually been applied mostly to industry data.

Computing Additive Risk Loads Instead of Risk Adjustments to the Discount Rate

Since the procedures described here focuses on computing a risk adjustment to the discount rate, the procedure to compute an additive, dollar-value risk load must convert the risk-adjusted rate into a risk load (as a ratio to the liability value). However, it is possible to compute the risk load directly using the same data for computing a risk adjustment to the discount rate. This approach might be preferred for a short tail line.

One approach to computing an additive risk load is simply to calculate the ratio of the profit on the policies at the beginning of the period to the average discounted losses, where losses are discounted at a risk-free rate rather than a risky rate. Thus, the risk load (expressed as a percentage of the present value losses) is equal to the present value of the premiums minus the present value of expenses minus the present value of the losses (plus loss adjustment expenses) divided by the present value of the losses. All quantities are discounted at the risk-free rate.

Unlike the adjustment to the discount rate, this risk load would not be meaningful unless computed by line, since the duration of the liabilities varies by line. An example of this computation is shown in the Appendix.

Method 5 - Actuarial Distribution-Based Risk Loads[35]

The evolution of this approach relative to pricing is given first, followed by the extension to the valuation of liabilities.

Pricing context

Probability-based actuarial risk loads are among the oldest procedures developed by actuaries for estimating the risk adjustment to losses. These approaches continue to develop, even as other approaches, which largely evolved from other disciplines (such as economics and finance), continue to add to the tools used for deriving risk loads. Distribution based loads arose in the context of insurance pricing to fill the perceived need to apportion the targeted underwriting profit to different classes of business according to their actual riskiness, as described mathematically by the probability distribution of the loss.

The first approaches to the problem focused on the volatility of the individual loss, characterized mainly by the severity distribution. In 1970, Hans Bühlmann set forth three possible principles that might be applied to the problem:

- The Standard Deviation Principle: Risk Load = λ SD[Loss],
- The Variance Principle: Risk Load = λ Var[Loss],
- The Utility Principle: U(Equity) = E[U(Equity + Premium - Loss)].

Actuarial distribution-based risk loads often invoke collective risk theory to explain the derivation of the risk load. Collective risk theory provides a model of the insurance loss generating process that can be used to derive aggregate probability distributions. The theory also allows derivation of the distribution parameters such as standard deviations or variances, which are used in the risk load formulas. Recent developments in collective risk theory have given rise to an additional principle used to derive risk loads:

- The expected policyholder deficit (EPD[36]) principle: Risk Load = λ Surplus Requirement.

Surplus is determined based on the expected policyholder deficit, which is derived from the

[35] This exposition draws heavily on Glenn Meyers' September 18, 1998 presentation to Casualty Actuaries of New England (CANE).

[36] The "expected policyholder deficit" is the total expected level of uncompensated losses over the total expected level of all losses, for a given level of assets (reserves plus surplus) supporting a risk. For example, assume 99% of the time losses are only $1, 1% of the time they are $100, and the total level of assets supporting this risk is $90. Then expected uncompensated losses are $0.10. Total expected losses are $1.99. The expected policyholder deficit is 0.10/1.99, or around 5%. For further discussion of this concept, see " Solvency Measurement for Property-Liability Risk-Based Capital Applications" By Robert P. Butsic, published in the 1992 CAS discussion paper program titled "Insurer Financial Solvency".

aggregate probability distribution of either losses or surplus (assets minus losses). This principle is very similar to the tail-value-at-risk principle proposed by Meyers.[37]

Each of the above principles contains an arbitrary coefficient λ, constant across classes of business (and concealed in the utility function), that can be adjusted to yield the desired overall underwriting profit or rate of return on surplus. In much of the literature the time element is not addressed explicitly. It is straightforward, however to apply the risk load to discounted liabilities.

The first two of the principles were applied in the practical context of increased limits ratemaking at the Insurance Services Office (ISO) in the late seventies and early eighties.

During the eighties, regulatory pressures brought the Capital Asset Pricing Model (CAPM) into the debate regarding how to incorporate risk into insurance prices. CAPM is founded on certain axioms that are violated in the context of insurance pricing (e.g., no default, frictionless markets), but this intrusion of modern financial theory stimulated much thought as to how the risk load formalism can address enterprise-wide and market-wide issues that had been neglected in the earlier formulations. The concept of systematic risk, already familiar to actuaries as parameter risk, was incorporated into practical treatments intended for actual insurance pricing.

The Competitive Market Equilibrium approach to risk load incorporates parameter uncertainty and other mechanisms, which generate correlations among distinct insurance contracts (e.g., the catastrophe mechanism, which can affect many contracts, in different lines of insurance, in a single event).[38] This scheme attempts to integrate capital market theory and collective risk theory in the development of risk loads for insurance pricing. The procedure requires all parties to agree that more variance is worse and less is better. (Note that the CAPM disagrees. It treats variance not related to the market as not valued by the market and not a concern, as it can be diversified away. It assumes no transaction cost to do so.)

The answer given by this scheme gives a contract risk loading proportional to the change in the variance of the insurer's bottom line caused by the addition of that one contract to the insurer's portfolio. This raised an interesting parallel with work being done at about the same time on reinsurance pricing based on marginal surplus requirements.[39] The Competitive Market Equilibrium result can be re-expressed in terms of the marginal surplus (risk capital) required to support the additional business, and thus linked to the cost of risk capital. More recent work using probability distributions has referenced the expected policyholder deficit concept, rather than standard deviation, variance or probability of ruin to motivate the computation of marginal

[37] Meyers, Glenn, "The Cost of Financing Insurance", paper presented to the NAIC's Insurance Securitization Working Group at the March 2000 NAIC quarterly meeting.
[38] Meyers, Glenn G., "The Competitive Market Equilibrium Risk Load Formula for Increased Limits Ratemaking," Proceedings of the Casualty Actuarial Society (PCAS), LXXVIII, 1991
[39] Kreps, Rodney E., "Reinsurer Risk Loads from Marginal Surplus Requirements," Proceedings of the Casualty Actuarial Society (PCAS), LXXVII, 1990, p. 196

surplus requirement and, therefore, of risk load.[40, 41]

Extension to Loss and Expense Reserves

The above methods apply prospectively to situations where the losses have not yet taken place and only rating information is available. For risk-adjusted valuation of insurance liabilities, such methods would apply to the Unearned Premium Reserve (UPR) and Incurred But Not Reported Reserves (IBNR). As long as one has some kind of runoff schedule giving estimates of number and type of claims not yet reported, one can apply these methods to estimate the variability of unreported claims.

Estimating the variability of reported claims is a different problem because of the information available to the insurance company about actual reported claims. Meyers has addressed the problem in the context of reserving for workers' compensation pensions, using a parametric model for the mortality table and calculating the variance of conditional future payments.[42] Hayne has used the collective risk model with information about claim counts and severities as the claim cohort ages and assumptions as to distributions and correlation structures to estimate the distribution of outstanding losses.[43] Heckman has applied distribution and regression techniques to estimating the expected ultimate value of claims already reported and of IBNR claims.[44] For the two latter methods, the conditional loss distribution provides the information needed to calculate risk loads for the reserves.

There are some unsolved problems associated with approaches based on probability distributions. Research is in progress to develop methods for measuring correlations of lines or segments of the business with other segments, but there is no generally accepted approach for incorporating correlations into the measure of risk. This is believed to be important, as these correlations may make a significant contribution to, and in some cases may reduce overall risk. In addition, some of the risk load procedures such as those based on standard deviation and variance approaches are not value additive. That is, the risk load of the sum is not equal to the sum of the risk loads.

Advantages

- Actuaries have used the approaches for a long time to compute risk loads.

---

[40] Meyers, Glenn, "The Cost of Financing Insurance", paper presented to the NAIC's Insurance Securitization Working Group at the March 2000 NAIC quarterly meeting.

[41] Philbrick, Stephen W., "Accounting for Risk Margins," Casualty Actuarial Society Forum, Spring 1994, Volume 1, pp. 1-87.

[42] Meyers, Glenn G., "Risk Theoretic Issues in Loss Reserving: The Case of Workers Compensation Pension Reserves," Proceedings of the Casualty Actuarial Society (PCAS), LXXVI, 1989, p. 171

[43] Hayne, Roger M., "Application of Collective Risk Theory to Estimate Variability in Loss Reserves," Proceedings of the Casualty Actuarial Society (PCAS), LXXVI, 1989, p. 77-110

[44] Heckman, Philip, "Seriatim, Claim Valuation from Detailed Process Models," paper presented at Casualty Loss Reserve Seminar, 1999.

- This is an area of active research with many worked out examples of how the method can be applied.
- The method is intuitive: risk load is related to actual risk for a body of liabilities.
- The data required to compute the risk loads is readily available within many insurance companies and many actuaries are qualified to perform the computation.
- Many reserving actuaries are familiar with using aggregate loss probabilities to establish confidence intervals around their reserve estimates.
- This method can be used with company-specific data.
- This method can be used by line to reflect unique line of business risks.

Disadvantages

- The approaches have often been criticized as being inconsistent with modern financial theory, as classically formulated, relative to compensation for diversifiable risk. For example, the risk loads often fail to satisfy the *one-price rule*, whereby two insurers offering identical insurance coverage would charge the same price.
- Sometimes the weight given to process risk relative to parameter risk in determining the risk load can appear to be too large. Many researchers and practitioners believe that risk loads apply only to nondiversifiable (parameter or systematic) risk not to unique (or process) risk. It should be noted that it is not universally accepted that only diversifiable risk matters when computing risk loads.[45,46]
- The risk loads may not satisfy value additivity. As a result, two companies with identical lines but a different mix can have different risk margins (see discussion below).
- A large number of methods for doing these calculations exist, yielding a variety of results. There is little guidance regarding which of the available methods is appropriate for a given set of circumstances.
- Certain parameters are not only subjective, but there is little guidance on how to calibrate them. For instance, only the more recent papers discuss a conceptual framework for selecting λ.
- Parameters are often determined in a subjective manner and may therefore be inaccurate.
- Actuaries are still struggling with measuring the correlations between lines of business. This may be a significant source of risk to companies.

Note that the lack of value additivity is not universally accepted as a disadvantage. For example, some believe there is much less risk in a $1 million (undiscounted) share of a large company's auto liability reserves than in the entire $1 million in undiscounted auto liability reserves for a small regional insurer. Thus, the former may be worth more than the latter (i.e., valued with a smaller risk margin).

---

[45] Cornell, Bradford, "Risk. Duration and Capital Budgeting: New Evidence on Some Old Questions", Journal of Business, 1999 vol 72, pp 183-200.
[46] Stulz, Rene, "Whats wrong with modern capital budgeting?", Address to the Eastern Finance Association, April, 1999

Method 6 - Using the reinsurance market to estimate the fair value of liabilities

The reinsurance market offers the most direct approach to estimating the fair value of an insurance company's liabilities. Blocks of liabilities are often sold either on a retrospective basis, in transactions such as loss portfolio transfers, or on a prospective basis in more commonly purchased excess of loss treaties. The price structures associated with these contracts provide another glimpse of the implicit risk load required to record the liabilities at their fair value.

Reinsurance prices may require some adjustment before they could be used to estimate the fair value of liabilities. For example, market prices offered by some reinsurers reflect an embedded option value equal to the value of their default on their liabilities. Such market prices would have to be adjusted upward to remove this default value. Another example is portfolio transfers that include customer lists or renewal rights. The effect of these lists or rights on the total price would have to be isolated and removed before the portfolio transfer price could be used for a fair value estimate.

There are numerous practical issues that need to be addressed before the method can be implemented in practice. For example, how would a ceding company measure the risk loading in the reinsurer's price structure? How could the analysis of a particular treaty structured to reinsure a portion of the company's liability be generalized to estimate the fair value of all its liabilities? Possible approaches are:

- Reinsurance Surveys: On a regular basis, leading companies can be surveyed to evaluate the risk loading implicit in their reinsurance structure. The survey can be structured to discriminate between various lines of insurance and sizes of ceding companies. The implicit risk loading can then be published and employed by all companies with a particular set of attributes (size, type of business, balance sheet leverage, etc.). Note that this is a controversial suggestion. (*Asking companies to share loss information is one thing. Asking them to share pricing information is something else entirely. First, the pricing "assumption" may not be as objective an item as a loss amount. It may be a gut call that varies by sale. Second, there are many more antitrust issues in sharing pricing information than in sharing loss information.*)

Conceptually, this would operate similarly to the PCS Catastrophe Options currently offered by the Chicago Board of Trade. These options are priced based on an index, which is constructed in the following way:

> *"A survey of companies, agents, and adjusters is one part of the estimating process. PCS conducts confidential surveys of at least 70% of the market based on premium-written market share. PCS then develops a composite of individual loss and claim estimates reported by these sources. Using both actual and projected claim figures, PCS extrapolates to a total industry estimate by comparing this information to market share data."* [47]

- Extrapolating from a company's own reinsurance program: Companies that submit their reinsurance programs to bid will receive reinsurance market price information from a number of providers. At a minimum, even the information contained in one well-documented bid may be sufficient to compare the reinsurer's price to the ceding company's best estimate of the ceded liabilities discounted at the risk-free rate. In practice, a number of adjustments to this risk load may be appropriate. For example, if the only reinsurance purchased is high layer excess, then the risk loading will be commensurate with the increased risk associated with that layer. Publicly available increased limits tables (e.g., ISO) might be suitable in some cases to evaluate the relative risk at each layer of coverage. An insurer's policy limits profile can then be employed to evaluate the weighted total limits of their liability portfolio and the resulting risk load.

Advantages

- The reinsurance market is the closest structure to a liquid market for insurance liabilities;
- Most insurers have access to the reinsurance market and can therefore gain information regarding their unique risk profile;
- Similar to catastrophe options, once the survey results are published, it would be relatively straightforward to estimate fair value

Disadvantages

- Results can be sensitive to capacity changes in the reinsurance market. As such, the values at any point in time may not represent future values. In fact, in highly competitive market cycles, a negative risk load could be obtained for some coverages.
- Unstable reinsurance prices also make it difficult to update estimates for each reporting period. If the information required for the fair value estimate could not be obtained quickly enough, all estimates would have to be recalculated each reporting period.
- The credit risk of the reinsurer's default on its obligation is embedded in the price. For reinsurance, this can be material, and would have to be removed, but the isolation of this item from the total price (and other risks) may be problematic.
- This approach would also raise difficulties in updating the values, as it would require

[47] Chicago Board of Trade web site: PCS Catastrophe Insurance Options – Frequently Asked Questions

regular surveys or continual shopping of ceded business to reset the risk charges.[48]

- Some reinsurance quotes are not transparent, so that the implied risk loading may be difficult to ascertain. Often, the insurer and reinsurer would each have different estimates of the expected loss and other components of price.
- The users of this method will only sample the reinsurance market. I.e., they will not be using the entire market for estimation. This could introduce bias.
- Reinsurance markets focus much more on prospective exposures rather than past exposures, partly due to current accounting treatment of most retroactive reinsurance contracts. As such, there are fewer market prices potentially available (and a much smaller market) for reinsurance of existing claim liabilities.
- Reinsurance prices embed antiselection bias. The price of reinsurance for the portion of an insurer's portfolio ceded may be higher than the price if all risks were ceded.

[48] Note that continual updates would be required under fair value accounting. This is because fair value accounting is meant to be an idealized market value, i.e., an actual market value if a sufficiently active market exists, or an estimate of what a fair market value would be otherwise. As such, a fair value estimate would have to be updated as often as an active market value would be updated. In general, market values in an active market change constantly.

Method 7 - Direct estimation of market values

This is the method of Allen, Cummins and Phillips.[49] In this approach, a time series of publicly traded insurer data is analyzed. The output of the analysis is an estimate of the market value of each insurer's liabilities for each year of the history. The market value of liabilities is derived by subtracting the market value of the equity from the market value of total assets. The market value of equity is calculated by extending the method of Ronn and Verma to avoid the problem of including intangible asset values in the equity measurement.[50] Here, the equity value is determined so that the measured volatility of the insurer's stock price and of its asset values are consistent. This method is described in the section on measurement of credit risk The market value of assets is estimated from the separate asset categories, most of which are publicly traded.

The market value of liabilities thus obtained contains an embedded option value equal to the value of default on the liabilities. This value of the default can be separately determined by the of Ronn-Verma method.

Adding back the default value gives the market value of the liability as if there were no credit risk. Next, the nominal (undiscounted) value of the liability is compared to the no-default market value to determine the implied interest rate at which the nominal value is discounted to get the market value. This calculation requires an estimation of the payment pattern of the liabilities (also used in the above-average payment duration). The risk margin, as a reduction to the risk-free rate, is the difference between the risk-free rate and the implied rate underlying the market value.

A numerical illustration of the method is shown in the Appendix.

Advantages

- The method is theoretically sound. It produces a risk load consistent with modern financial theory without requiring the calculation of a beta.
- The method is objective and the analysis is reproducible
- The method is a type of direct measurement of liabilities that may be desirable by the accounting profession. However, the measurement is direct for the industry, but not for a particular company

Disadvantages

- There are difficulties with the estimation of parameters:
  a) Some insurers in the data are also life insurers, or involved in multiple lines not relevant to a particular company at issue; carving them out requires estimating the

[49] Allen, Franklin, J. David Cummins and Richard D. Phillips, 1998, "Financial Pricing of Insurance in a Multiple Line Insurance Company", Journal of Risk and Insurance, 1998, volume 65, pp. 597-636.
[50] Ronn, Ehun I., and Avinash K. Verma, 1986, Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model, Journal of Finance, 41(4): 871-895.

market equity value of these other operations.

b) Some companies are members of financial conglomerates, or general conglomerates (e.g., General Electric).

c) Not all insurers are publicly traded. These include foreign companies, privately held companies and mutuals or reciprocals.

- The liabilities may be under- or overstated in the financial statements. Therefore, the market value may reflect an adjustment to the book value, based on market perceptive of this bias. Any perceived change in this bias may make prior history unusable.
- Measurement problems make it difficult to provide a stable estimate for individual line of business risk margins. It is also difficult to get a reliable estimate for an individual firm.
- Most actuaries don't have any experience with this method. It has not yet been used in practice.

Method 8 - Distribution Transform Method

A number of authors have proposed risk-loading procedures based on transforming the aggregate loss probability distribution.[51] The risk-loaded losses are computed from the mean of the transformed distribution. A simple example of such a transform is the scale transform:

$$x \rightarrow kx$$

where x = the aggregate losses
$k > 1$

As a simple, but unrealistic example (because insurance losses tend to have positive skewness), x is a normal variable, that is, if aggregate losses follow a normal distribution and k is 1.1, then the loss distribution's expected mean is shifted upwards by 10%. Thus, a company purchasing the liabilities would require 10% above the present value of the liabilities (at a risk-free rate), in order to be adequately compensated for the riskiness of the liabilities. If one is using this distribution to compute primary losses for an exposure where the limits applied to losses in the aggregate, the expected mean would be increased by less than 10%, but losses excess of the primary limit will be increased by more than 10%.

In the more recent literature on the transform method the power transform is used.[52] (Other transforms such as the Esscher transform also appear in the literature). This approach raises the survival or tail probability to a power.

$$S^*(x) = S(x)^r$$

where $S(x)$ = the original survival distribution, $1-F(x)$, or 1 minus the cumulative probability distribution);
$S^*(x)$ = the transformed survival probability.

If r is between 0 and one, the tail probabilities will increase and the transformed distribution will have a higher mean than the original distribution.

The choice of the transformation parameter r is guided by the uncertainty of the business being

[51] Venter, Gary G., 1991, Premium Implications of Reinsurance Without Arbitrage, ASTIN Bulletin, 21 No. 2: 223-232. Also,
Wang, Shaun, 1998, Implementation of the PH-Transform in Ratemaking, [Presented at the Fall, 1998 meeting of the Casualty Actuarial Society]. Also,
Butsic, Robert P, 1999, Capital Allocation for Property Liability Insurers: A Catastrophe Reinsurance Application. Casualty Actuarial Society Forum, Fall 1999.
[52] Wang, Shaun, 1998, Implementation of the PH-Transform in Ratemaking, [Presented at the Fall, 1998 meeting of the Casualty Actuarial Society]. Also,
Venter, Gary G., 1998, (Discusssion of) Implementation of the PH-Transform in Ratemaking, [by Shaun Wang; presented at the Fall, 1998 meeting of the Casualty Actuarial Society]

priced. The greater the uncertainty, the lower r will be. In practice, this may mean that one calibrates the parameter by selecting a transformation that approximates current market premiums for a given class of exposures. Wang suggests that using a distribution transformation to derive risk loads is the equivalent of including a provision for parameter risk, but not process risk, into the formula for risk loads. Thus, one might select r based on subjective probabilities about the parameter uncertainty of the business.

Wang (1998) has suggested that one could apply this approach in two ways.[53] The first applies a transform separately to the frequency and severity distributions used to price policies. The second transforms the probability distribution of aggregate losses (i.e., the convolution of the frequency and severity distributions). However, Venter suggests that one could obtain inconsistent results when applying a transform to aggregate losses, and prefers working with the frequency and severity distributions.[54]

Option pricing theory and the distribution transform method are related. The parameters of the probability distributions used in the option pricing formulas typically reflect "risk neutral" probabilities, rather than real probabilities. Thus, for example, the parameters used to price interest rate options are generally derived from current actual prices of bonds of different maturities, or from the current yield curve, rather than from empirical time series data of the various interest rates. One could view the "risk neutral" probabilities as a transformation of the distribution for the underlying asset values.

Advantages

- The method produces a risk load consistent with modern financial theory without requiring the calculation of a beta. Risk loads are value additive. (Note again that there is not universal agreement among actuaries that risk loads should be value additive.) The approach is similar to that used in pricing options.
- The method is conceptually straightforward to understand and explain. Once r or a similar parameter has been selected, it can be reused subsequently.
- This approach is currently used in reinsurance pricing.
- It is theoretically viable for estimating risk loads by layer. Many of the other methods do not address layers or deductibles.
- It is an area of active research for those investigating risk load methodologies.

Disadvantages

- It is not in common use for producing prices or risk loads on primary business. Currently its primary use is in producing risk load for layers.

---

[53] Wang, Shaun, 1998, Implementation of the PH-Transform in Ratemaking, [Presented at the Fall, 1998 meeting of the Casualty Actuarial Society].

[54] Venter, Gary G., 1998, (Discusssion of) Implementation of the PH-Transform in Ratemaking, [by Shaun Wang; presented at the Fall, 1998 meeting of the Casualty Actuarial Society]

- As currently applied, in order to calibrate the parameters, it often requires knowledge of the risk loads on primary business.
- Because it is a new approach, actuaries are not as familiar with it as with some of the others presented in this paper.
- The parameters may be selected based on the analyst's experience with a particular line of business. This introduces an element of subjectivity, where different analysts may choose different values for the parameter.
- It is not clear which transform choice to use. Many of the transformation methods are chosen for their mathematical tractability, and are not supported with empirical evidence.

Method 9 - The Rule-of-Thumb Method

The methods presented so far require that the person computing the risk-adjusted present value of liabilities do original analytical work. In some situations there may not be adequate data or other resources to develop the risk adjustment from scratch. In such situations it might be appropriate to use a rule of thumb that provides a "quick and dirty" way to derive a risk adjustment. Such methods would be relatively easy to apply but would produce broadly reasonable results. Examples of rules of thumb would be:

- Compute a risk adjusted discount rate by subtracting 3% from the risk-free rate.
- The risk load should be 10% of the present value of General Liability liabilities and 5% of the present value of Homeowners liabilities.

The numbers in the examples above are for illustrative purposes only. A separate body of actuaries and other experts could determine actual guideline values. This group would review existing research and perform additional studies where necessary. Quite likely, it would consolidate results from using one or more of the other methods in this document.

Advantages

- For the individual company, it would be simpler to apply than any of the other alternatives. It would reduce the work effort for actuaries and others, who would not have to separately develop risk adjustments.
- This approach may lead to industry standard risk adjustments being used, thus creating comparability from company to company.
- It may reduce the likelihood that a risk adjustment methodology can be used to manipulate a company's financial statements.

Disadvantages

- Fair values produced using this approach may be less accurate because the unique risk factors for a company may not be reflected.
- It precludes actuaries from applying methods that reflect new developments for determining risk adjustments.
- An industry body may be required to perform research to parameterize the risk adjustments. This may create antitrust issues. It is not clear that the industry body would be sufficiently authoritative for its research to be used in financial valuations.

Method 10 - Alternative Methods

This paper has presented a number of possible approaches to estimating the fair value of insurance liabilities. Most of these approaches are rooted in analytical methods documented in the actuarial literature. However, research continues into how to determine risk adjustments. Not all current developments are covered in this paper and undoubtedly others will be published. A company may wish to use alternative approaches not presented in this paper. In such cases, there are a number of points one should consider:

- Once selected, the approaches should be used consistently. Changing approaches from year to year may result in inappropriate income statements.
- If the method is changed, it should be documented adequately.
- The risk margin should be positive.

Converting a risk adjusted discount to an additive risk load

A number of the methods presented in this paper produce an adjustment to the risk-free discount rate. Risk adjusted present values of liabilities are then derived by discounted the liabilities using the risk-adjusted rate. An approach to deriving a dollar-value risk load is to work from the risk-adjusted discount rates. This approach might be used if one wanted to discount losses at the risk-free rate and apply the risk load to the losses directly. The procedure begins by discounting the liabilities at the risk-adjusted and the risk-free rate. It then computes the difference between the two discounted quantities. The risk load is this difference divided by the present value of the liabilities, discounted at the risk-free rate. The table below presents an example where this calculation is performed for liabilities of various durations, when the assumed risk-free rate and the risk adjustment remain constant.

Risk Free Rate:	6.0%		
Risk Adjustment	3.0%		
Duration	PV @ Risk-Free Rate	PV @ Risk-Adjusted Rate	Risk Load
1	94.3%	97.1%	2.8%
2	89.0%	94.3%	5.6%
3	84.0%	91.5%	8.3%
4	79.2%	88.8%	10.8%
5	74.7%	86.3%	13.4%
6	70.5%	83.7%	15.8%
7	66.5%	81.3%	18.2%
8	62.7%	78.9%	20.5%
9	59.2%	76.6%	22.8%
10	55.8%	74.4%	25.0%

Unearned Premium (or Unexpired Policy) liability methods

As noted in the background section, a fair value accounting system focuses on the measurement of assets and liabilities, not income. As such, the current recording of unearned premium under U.S. GAAP accounting conventions would be replaced with the fair value of the business written but not yet earned. The methods used to estimate this fair value have much in common with the above methods that estimate the fair value of the liabilities for unpaid losses. However, additional methods may be applicable since it may be easier to discern the market prices underlying earned premium. One can argue that the booked premium represents the "market price" charged by the particular insurer.

One area where such additional methods may be needed is property insurance, particularly where catastrophe exposure exists.

Possible methods to consider include:

- The price at which the business was written, the original entry price. The initial fair value for a policy's liability may be the premium charged (less expenses).
- The price at which the company is currently writing similar business.
- The price at which similar business is currently being written by the market, e.g., a broad average price. It is an indication of the current entry price. (This value may only be available retrospectively shortly after the balance sheet date.)
- The price at which reinsurance is being purchased for this risk, both quota share reinsurance, which prices the entire risk, or excess of loss reinsurance, which should provide a market guide to one of the more volatile components of the risk. This also is an indication of the current exit price.
- An actuarial estimate of the expected value of discounted losses associated with the business written but not yet earned, adjusted for risk. The estimate of the necessary risk adjustment would be based on the above methods for estimating the market value of unpaid losses. In particular, return on equity models, internal rate of return models, and models based on the aggregate probability distribution of losses, can be directly applied to future losses (losses not yet incurred on business written).

Note that the actuarial methods applicable to lines of business that contain a significant catastrophe potential may require modification to consider the seasonality of the exposures.

Summary

A number of methods for computing risk adjustments to discounted liabilities have been presented. These are the approaches that the committee thought were worthy of discussion. Not all would be feasible for the individual company actuary to implement. As fair value becomes established as an accounting procedure, more research and application will be performed, and more methods will become feasible.

Some methods would require an "official" body such as a committee of the American Academy of Actuaries to perform research to establish parameters. Once established, the parameters could thereafter be used at individual companies without further research or analysis being required. This would hold only if one agrees that it is acceptable to ignore risks that are unique to companies, such as those classified under diversifiable risk.

Methods such as those based on CAPM and IRR pricing models should be straightforward to modify for estimating the fair value of liabilities. Actuaries are also well acquainted with methods based on aggregate probability distributions. Actuaries should be able to apply one or more of the methods to a line of business for which they are computing risk-adjusted discounted reserves.

Some methods are more appropriate for some lines of business. For instance, methods based on using risk-adjusted discount rates have been applied to lines of business with longer tails such as Automobile Liability and Workers Compensation. However, they may be inappropriate for short tail volatile lines such as property catastrophe because the risk is not time-dependent. Methods based on applying aggregate probability distributions might be appropriate for such short tail volatile lines. However, their use outside of increased limits and catastrophe pricing has not been well researched.

The direct estimation method is relatively new and has only been applied by academic researchers. Therefore, it could be difficult for practitioners to apply until further study has been done. Using reinsurance pricing to develop a risk load is, in principle, the most consistent with computing market-based estimates of liabilities. However, due to limitations on available data, the extent of the market and a lack of published research on the approach, it might be difficult to apply in practice. There might be special situations where it could be used, such as in evaluating catastrophe liabilities.

In general, risk adjustments based on industry-wide information will be more stable than risk adjustments based entirely on company-specific data. Also, risk adjustments based on individual line of business data will be less stable than risk adjustments established using all-lines data. However, such risk adjustments will fail to incorporate some of the risk components of that are unique to lines of business or to companies.

This summary and discussion provided by the task force of methods available for computing the risk adjusted present value of liabilities demonstrates that actuaries have the theoretical understanding needed to implement fair valuing of insurance liabilities. We have identified a number of models that are available and appropriate for actuaries to use in estimating fair value liabilities. No issues have been identified that are not susceptible of actuarial estimation.

The following table summarizes our findings on the methods of deriving risk adjustments.

Summary of Features of Estimation Methods

Method	Uses Industry Data	Uses Company Specific Data	Has Specific Time Element	Uses Leverage Ratios	Incorporates Systematic Risk	Incorporates Process Risk	Is Value Additive	Commonly Used in Pricing	Commonly Used for Reserve Margins
CAPM	X		X	X	X		X		
Internal Rate of Return	X	X	X	X	X		X	X	
Single Period RAD	X	X	X	X	X		X	X	
Using Under-writing Results	X		X		X	X	X		
Based on Probability Distributions	X	X			X	X			X
Based on Reinsurance	x	x							
Direct Estimation	x		x		x		x		
Distribution Transforms	x			x	x	x			
Naive Methods	x				x		x		

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section E - Accounting Presentation Issues

The purpose of this section is to discuss financial reporting presentation issues resulting from a change to fair value accounting. Financial reporting presentation deals with the design of the reporting template, i.e., what financial values should be displayed, and in what format. It assumes that any required value can be determined, such as through the various methods in Section D. While many implementation issues may arise from the choice of a particular reporting template, such issues will not be discussed in this section. All implementation issues will be discussed in the next section (Section F), whether arising from the estimation method chosen (Section D), or arising from the presentation template chosen.

The following actuarial presentation issues will be discussed. This list is meant to stimulate awareness of the various actuarial issues/concerns surrounding presentation and fair value accounting. It is not meant to give definitive guidance on how presentation should be done. The final choice of any presentation template is a judgment call, depending on the goals, priorities and preferences of the template designer(s).

- Historical loss development:
  - risk margins
  - time value of money
- Disclosure of fair value estimation methods.
- Gross versus net (of reinsurance, other recoverables).
- Recognition of premium revenue.
- Income classification:
  - Unwinding of interest discount
  - Interest rate changes
  - Experience adjustments, changes in assumptions.
- Consistent treatment of assets and liabilities
- Different financial statements for insurance vs. noninsurance entities
- Disclosure of credit standing impact
- Consolidated financial statements
- Regulation and tax requirements

1. *Historical Loss Development* - Currently some financial statement exhibits show historical loss development. These exhibits are useful for evaluating management's previous estimates of liabilities, and for evaluating the risk inherent in the estimates. Should these exhibits show historic fair value estimates? Issues associated with doing so on such exhibits include:

   a) Risk margins. The risk margin for a given coverage year runs off over time to a value of zero as the losses are paid. In addition, the perception of risk changes over time. For example, the risk margin of hurricane losses would have been valued less before the recent large hurricane losses in Florida. The perceived risk for mass tort liabilities is also now much greater than believed in the 1970s and prior. Are the purposes of these historical exhibits furthered or distracted by including historic risk margin estimates in the reported history?

b) Time value of money. The amount of discount runs off to zero as losses are paid out. Interest rates also fluctuate over time. As such, historical exhibits that reflect the time value of money might show development trends impacted strictly by changes in new money investment yields or the unwinding of interest discount. The economic impact of these trends depends on the how the corresponding asset portfolio was impacted. How should the historic loss development exhibits handle this issue?

A possible way of addressing the above two sub-issues might be to require historic loss development exhibits to be on an undiscounted, expected value basis. This would isolate the issues surrounding the expected value estimate (although it would ignore the issues surrounding the amount of the discount or risk margin). An alternative approach for evaluating the amount of the discount would be to require loss development exhibits to show all actual and projected values discounted back to the beginning of the coverage year. This would allow reflection of time value of money issues and expected value estimate issues, without the distortion from interest rate fluctuations. The issue would remain regarding whether to use the historical interest rates at the first valuation of the coverage year or restate at the current interest rates.

2. *Disclosure of fair value estimation methods* - Should the methods used to determine the fair value estimates be disclosed in the financial statements and if disclosed, where, and in what levels of detail? Depending upon the method(s) employed, the fair value components may differ by line of business as well as subline of business, duration of payments, location of the liabilities, and the currency that will pay out the liabilities. In addition, any changes to the method(s) or the values used to determine the fair value of liabilities may need to be disclosed in the financial statement.

3. *Gross versus net (of reinsurance, other recoverables).* - A decision needs to be made with regard to how much of the fair value presentation should be on a gross versus net basis. Should fair value adjustments be included in both gross and recoverable reportings, or would an overall net adjustment suffice? Where various amounts are reported in more detail, should these fair value adjustments be disclosed in the aggregate or by individual reinsurer or excess insurer (for a self-insured's financial statement)?

4. *Recognition of Premium Revenue* - How should premium revenue be recognized, under a fair value accounting system? Currently, premium revenue is recognized for property/casualty companies based on earned premiums, which equal written premiums plus the change in the unearned premium reserve. Since fair value accounting would require estimating the future losses associated with the unexpired portion of the policy, should this estimate of future losses be included in the loss reserves, and premium revenue become written premium? If so, the unearned premium reserves could disappear.

Long duration policies cause additional presentation issues if premium revenue is defined as written premium. Should revenue from long duration policies be reported or disclosed separately in financial reports, so as not to distort analyses of annual exposure growth? These policies may also distort otherwise reported policy year loss development trends. Should a single long duration policy be broken into separate 12-month policies for the purposes of policy year loss development exhibits?

Special policy features such as death, disability, and retirement benefits may also be impacted by a change in premium recognition. Should such benefits be accounted for as loss reserves or as unexpired policy benefits, under a fair value system?

5. *Income classification.* - Under a fair value accounting system, recorded balances (such as loss reserves) will reflect the time value of money, estimated future cash flows, and risk adjustments. Any of these components are subject to change over time, as the balance runs off. How should the changes in this components be reflected in income? The following discussion contains a discussion of the components.

a) Unwinding of interest discount – The principal question here is whether the unwinding of interest discount should be separately reported in income, and if so, where? Currently when companies discount property/casualty loss reserves for anticipated investment income, the unwinding of this discount over time flows through underwriting income, as a change in incurred losses, and is not separately identified. Discount unwinding for life insurance reserves also flows through as a change in incurred losses, but is separately identified in U.S. statutory accounting statements. Alternatively, the unwinding could be reported as interest expense, not in underwriting income.

Reflection of this unwinding in incurred losses maintains consistent treatment of any item affecting paid or outstanding losses, at the cost of distorting comparisons of losses to charged premiums. This distortion is caused by premiums being fixed in time, with no reflection of future investment income potential. If loss reserve discount is all unwound in incurred losses, then reported histories of incurred losses to premiums will tend to show excessive loss ratios for any long-tail line, distorting the true profitability picture. Reflection in interest expense allows more direct comparisons of losses to charged premiums.

b) Interest rate changes. How should changes in market interest rates used in discounting existing liabilities be reflected? Should the effect of these changes flow through underwriting? Should the effect flow through investment earnings? Should it be reflected in the same manner as unrealized capital gains, as a change in interest rates should affect both liabilities and assets similarly in a matched portfolio? Or should changes in loss reserves for any purpose other than unwinding of discount (e.g., change in expected ultimate payout, change in expected payment pattern, change in interest rates, etc.) all be reported in the aggregate, with no differentiation as to the cause?

c) Experience adjustments, Changes in assumptions. Another issue is how should an insurer present the effect of experience adjustments and changes in assumptions? Should changes due to actual cash flows being different from expected be reported separately from changes in assumptions about the future? The first are "realized" and the second are currently "unrealized". Should there be an effort to keep consistency with how similar issues for invested assets are treated? Should changes in risk margins be isolated, or combined with changes in any other assumptions?

6. Consistent Treatment of Assets and Liabilities - This issue arises whenever recoveries are available (beyond the initial premium) to offset changes in the estimated liabilities. Examples include retrospectively rated insurance policies, deductible policies, policyholder dividends, (re)insurance policies for which reinsurance (or retrocession) protection exists, and contingent commission plans (on reinsurance contracts). In these examples the change in a claim (or similar) liability should lead to an offsetting change (either in full or partial) in either an asset or another liability.

   For example, a direct retrospectively rated insurance policy may be subject to reinsurance. This could result in at least three balance sheet entries after losses have started to occur:
   - a liability for direct claims
   - an asset (liability) for additional (return) premiums on the retrospectively rated policy
   - an asset or contra-liability for the portion of the claim liability that is recoverable from reinsurers.

   The presentation issue regards the manner of reporting these amounts and their fair value adjustments in a consistent manner, and in such a way that their individual adjustments will not easily be taken out of context.

   (Note that to the extent the retrospective rating plan and the reinsurance coverage transfer risk, the overall net risk adjustment for all three items should be less than the risk adjustment on direct claim liabilities. This implies that the risk adjustment for some of the individual components may be a help to surplus.)

7. Different Financial Statements for Insurance Versus Non-Insurance Entities - Should financial statement requirements differ for insurance versus non-insurance entities? This issue arises when comparisons are attempted between insurers and self-insurers, traditional insurers and captive insurers, or insurers and other financial services companies selling similar products. The issue also arises with consolidated financial statements when the reporting entity includes both insurance and non-insurance operations.

8. Disclosure of Credit Standing Impact – If the fair value of liabilities is to include the impact of credit standing, these impacts should probably be disclosed separately in the financial statements. (The credit standing issue is discussed in more detail in Section H.)

9. Consolidated Financial Statements – Fair valuation generally requires that transactions be measured as if they were at arms-length. A key question regarding consolidated versus legal entity reporting is the difficulty in measuring fair value for legal entities of the same quota share group, especially when applied to a fresh start valuation of old claim liabilities. Thus, it may be necessary to estimate fair value for each pool member's direct book of business separately, rather than determining the fair value of the total quota share pool and then allocating the total pool result to the pool members.

   A related issue is how to report values containing risk margins if the component reporting entities have risk margins that do not add to the total risk margin of the consolidated entity. Should the component risk margins be scaled back to show value additivity?

10. Regulation and Tax Requirements – The change to fair value will impact both the absolute value of many of the statement items as well as the format of the financial statements. This may impact existing regulatory and tax use of financial information that may have come to depend on the existing financial statements. The final "fair value" statements may have to include accommodations for these needs. Alternatively, the regulatory and tax processes could be changed to adapt to the new financial statements. A third alternative would be to create additional supplemental reporting, based on the old accounting standards, as if nothing had changed. Examples of areas potentially impacted include federal income taxes, solvency testing, and market conduct exams.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section F - Implementation Issues

Introduction

Up to now, this paper has dealt primarily with two areas associated with fair valuing insurance liabilities. The first of these areas was contained in Sections C and D, "Fair Value Alternatives" and "Methods of Estimating Risk Adjustments." Sections C and D discuss a variety of ways that a liability's fair value could be determined in theory. The second area addressed so far in the paper was that of presentation. This was the subject of Section E.

The current section, Section F, goes the next step, discussing issues arising from the implementation of these concepts and methods and presentations. Implementation issues can be categorized as:

1. **Issues related to the availability and usability of market information.** These include:
   - 1.1. The robustness of the transactions occurring in the marketplace.
   - 1.2. Intangibles included in market prices that might not be relevant in a fair value liability valuation.
   - 1.3. Influence of information asymmetry on market prices.
   - 1.4. The existence of disequilibriums or temporary disruptions in market prices.
   - 1.5. The lag between event occurrence and the reporting of the event in the marketplace.
2. **General issues related to developing parameters for fair value methods.** These are issues that are not related to any particular fair value methodology. Rather they deal with concepts that can be thought of as some of the theoretical underpinnings of fair value accounting. These include:
   - 2.1. Whether or not a risk charge should always be included in the fair value of a liability.
   - 2.2. What properties a risk charge should have, specifically related to the inclusion of a value for diversifiable risk and value additivity.
   - 2.3. Whether an adjustment for an entity's own credit risk should be included in that entity's fair value valuation of its liabilities.
   - 2.4. The issues that need to be weighted when deciding to use industry-wide data or company-specific data in a fair value calculation.
3. **Application of fair value methodologies – general issues.** This section discusses issues that relate to questions that fair value practitioners will need to address when preparing fair value financial statements. These issues are ones that relate to how to physically create numbers to put on fair value financial statements, but that are not specific to any one methodology. Included under this heading are:
   - 3.1. The steps the actuarial profession might need to take to prepare for the implementation of a new requirement.

3.2. What items should contain fair value adjustments in their carrying value?

3.3. How renewal business ought to be considered when developing fair values.

3.4. How judgment should be accommodated when developing fair value estimates.

4. **Application of fair value methodologies — method-specific issues.** This section highlights issues associated with different methods that a practitioner ought to be aware of when choosing a fair value methodology. The specific issues being highlighted are:

   4.1. Methods that rely on CAPM.

   4.2. Methods that rely on public data.

   4.3. Methods that produce results on a total company-basis only.

   4.4. Time period sensitivity of some methods.

   4.5. The inclusion or exclusion of a value for process risk in valuations created by different methods.

   4.6. The existence or lack thereof of value additivity in valuations created by different methods.

   4.7. The appropriateness of different methods for the valuation of volatile, short-tailed lines of business.

5. **Presentation issues**. These are issues associated with the actual presentation of results in a fair value financial statement. Items include:

   5.1. Updating carried values from valuation date to valuation date, especially between full-scale analytical re-estimations of appropriate carrying values (in accounting parlance, a "fresh-start" valuation).

   5.2. Issues associated with the initial development of exhibits that show historical development.

1. Issues related to the availability and usability of market information

This is the first item to be discussed because it is FASB's and the IASC's stated preference that market valuations be used wherever possible. However, we are skeptical as to the usability of market information for developing fair value valuations of insurance liabilities. The five specific reasons for this skepticism are as follows:

1.1. Is the observed market active and robust enough for fair value estimation purposes? A key principle espoused by both FASB and the IASC is that the first choice for the development of fair values is from the marketplace.[55] However, there is not currently much of an active market that can be used to establish price comparisons. Moreover, the transactions that are being done may suffer from a lack of "market relevancy" whereby the marketplace transaction was for a block of liabilities that was similar but not exactly the same as the block of liabilities a company is trying to value. The company in this situation is faced with trying to decide how the market would respond to the differences between the company's liabilities and those that were involved the marketplace transaction.

1.2. The observed market values may contain intangibles not relevant to the valuation at hand. A similar but unrelated marketplace issue is the quantification of the value of noneconomic considerations in a market price. A company could have a variety of reasons for accepting one market price over another that are particular to that company. One example could be the nature of the relationship that exists with a particular reinsurer. The chosen reinsurer might not be the lowest cost option available to the company, but because the company trusts its relationship with the reinsurer, the company may feel the noneconomic "relationship value" is worth the extra cost. A different company looking to price a similar block of liabilities might not have the same relationship with a reinsurer. For the second company, then, the relationship value does not exist and the market price assigned to the first company's liabilities would not be appropriate valuation for the second company's liabilities.

[55] There is no universally accepted definition of "fair value" to-date, although they all follow the same general concept given by this short definition. The detailed definition that FASB is proposing can be found in FASB's Preliminary Views document titled "Reporting Financial Instruments and Certain Related Assets and Liabilities at Fair Value," dated December 14, 1999, and labeled "No. 204-B." The definition starts on paragraph 47, with discussion and clarification continuing through paragraph 83. Paragraph 47 states: *"Fair value is an estimate of the price an entity would have realized if it had sold an asset or paid if it had been relieved of a liability on the reporting date in an arm's-length exchange motivated by normal business considerations. That is, it is an estimate of an exit price determined by market interactions."*

The IASC has a similar definition (found on page A181 of their Insurance Issues Paper, released November 1999). It reads: *"The amount for which an asset could be exchanged, or a liability settled, between knowledgeable, willing parties in an arm's length transaction."*

1.3. Available market information, such as stock analyst estimates, or isolated reinsurance prices may not be reliable due to information asymmetry. The market price for an actual liability traded on an active market is likely to be quite different than the market value of an insurer's entire portfolio of liabilities. It is the latter item that is important in fair value accounting, not the former. Unless all the insurer's liabilities are transferred, the assuming reinsurers will quite rationally believe that the ceding insurer is selecting against the reinsurer. This situation arises because the market (the reinsurers) does not have access to the insurer's private information on the liabilities. Thus, the "actual market price" might not be a better fair value representation than an internal cash flow-based measurement unless most of the insurer's liabilities are actually transferred.

1.4. Market data available at a given valuation date may be distorted by disequilibriums or temporary disruptions. The existence of an underwriting cycle can be viewed as tangible evidence of the ongoing disequilibrium in the insurance marketplace, whereby product pricing swings back and forth between underpricing and overpricing generally over a seven-to-ten-year cycle. Market disruptions can be characterized as new events that lead to significant uncertainty and temporary disruption in the market for insurance products. Examples can include a threatening hurricane, a newly released wide-ranging court decision and new legislation (e.g., Superfund, or California Proposition 103). At such times, market prices right after the event may be wildly speculative, or the market may even be suspended, greatly complicating the use of market prices for fair value valuations.

1.5. The data available in the marketplace may be out of date. Depending upon the source being considered, there are often lags between event occurrence and event reporting. For example, an insurer, on behalf of its participation in an underwriting pool, may be exposed to certain liabilities that will ultimately be shared by all members of the underwriting pool. If someone were to base a fair value estimate on the pool's reported financials, the fair value estimate could reflect a lag of anywhere from several months to several years between when the pool actually experienced the results being reported and the reporting of them.

2. General issues related to the development of parameters for fair value methods

These issues are ones that do not specifically pertain to any one fair value method. These are "concept-type" items. Some of these, such as risk charge and credit risk, are items that relate to the general concepts that will underlie fair value implementation. Others, such as the use of industry-wide versus company-specific assumptions are issues that can not be resolved with a global decision and instead will need to be considered each time that a fair value methodology is applied.

2.1. Should a risk charge always be incorporated into the fair value of a liability? Most of the guidance to date (from the FASB and IASC) mandates including such a risk charge when it is material and estimable, and can be "estimated" from market information.

Paragraph 62 of FASB's Statement of Financial Accounting Concepts No. 7, Using Cash Flow Information and Present Value in Accounting Measurements, says:

> *"An arbitrary adjustment for risk, or one that cannot be evaluated by comparison to marketplace information, introduces an unjustified bias into the measurement. ... in many cases a reliable estimate of the market risk premium may not be obtainable.... In such situations, the present value of expected cash flows, discounted at a risk-free rate of interest, may be the best available estimate of fair value in the circumstances."*

Given that there is no active market for many insurance liabilities, there is no readily available, direct information on the market risk premium associated with their fair value. The market risk premium would have to be estimated. It is unclear as to what marketplace information would be required under such guidance for an acceptable estimate of the risk premium. Would the information have to be insurance specific or even insurance product specific, or could it be based on overall market pricing for risk in general financial markets? It is also unclear how much judgment may be used to produce an acceptable estimate of this risk premium.

If the guidance is worded and interpreted too stringently, then it may never be possible to include a risk premium in the fair value of insurance liabilities. Liabilities of high risk would be indistinguishable from liabilities of low risk, as long as the present value of expected cash flows was the same. More lenient interpretations may allow risk premiums for the more common liabilities, but the more unusual or higher risk liabilities may not qualify for a risk premium. This would result in a lower liability value (due to absence of a risk premium) for the highest risk items, a counterintuitive result. Attempts to always include a risk margin may raise reliability and auditability issues.

2.2. What properties should risk margins have? The following two items are separate, but related. They are separate in that each is an issue in its own right, but they are related in that it may only be possible to reflect one or the other, depending upon the fair value methodology that is chosen. For example, a methodology that reflects process risk in each line of business within a company might result in a series of fair values for each line, that when added together, produce a fair value in excess of the fair value that would be applicable to the company as a whole. This would be a reflection of process risk that violates value additivity. Both of these are discussed in greater detail in Section D.

- Should a value be placed on process (diversifiable) risk in the valuation?
- Should results have value additivity or not?

2.3. Should an adjustment for an entity's credit risk be incorporated into that entity's fair value of its liabilities? Section H contains the discussion of this issue.

2.4. **Use of industry-wide assumptions.** The two options for data and assumptions to be used in the methodologies described in Sections C and D are industry-wide or company-specific ones. Consideration must be given to the balance between the greater reliability of the industry data and the greater applicability of the company-specific data. Availability of data at the industry or company level is also a factor in selecting data for risk adjustment computations. Industry-wide data provides more consistent and reliable results, but may overlook important differences between the risks underlying the industry data and the company-specific risks being valued. Company-specific data will be more reflective of the underlying nature of the risks being valued, but the volume and the volatility of the data must be considered. If the company-specific data is too sparse or too volatile, it might not be usable. This is an issue that will need to be addressed on a situation by situation basis.

3. Application of fair value methodologies – general issues

These issues relate to the questions that fair value practitioners will need to address when preparing fair value financial statements. These issues are ones that relate to how to physically create numbers to put on fair value financial statements, rather than concept issues such as "what does fair value mean?" In this section, application issues are divided into two groups: issues that are not specific to any one methodology ("general" application issues) and those that are methodology specific. This segment will address the general issues. The methodology-specific issues are discussed after the general issue discussion.

3.1. **What steps will the actuarial profession need to take to prepare for the implementation of a new requirement?** As with any new requirement, the switch to a fair value valuation standard for property/casualty insurance liabilities would probably result in many unanticipated consequences. Many of these consequences would not be evident at first, and may take time to resolve once they are discovered. This may involve refinement of existing and development of new actuarial models and revisions to the initial accounting standards.

3.2. **Fair value accounting will affect more than just loss reserves. Should the same methodologies that are being used for loss reserves also be used for other items? How can consistency of underlying assumptions be maintained in the valuation of all items with fair value adjustments?**

Examples of the items that might warrant fair value adjustments include:

- The liability associated with the unexpired portion of policies in-force at the valuation date
- Liability associated with the unexpired portion of multi-year contracts
- Reinsurance contracts with embedded options, including commutation terms, cancellation terms, contingent commission provisions, etc.
- Differences between the fair value of liabilities on a net basis versus a gross basis
- Accrued retrospective premium asset or liability

- **Salvage and subrogation**

The real issue is not so much ***what*** contains fair value adjustments as ***how*** the adjustments are to be made. The accounting standards will determine those items that should contain fair value adjustments. The challenge will be to quantify the adjustments for these different items in a manner that is consistent with the adjustments underlying loss reserves. The implementation issue facing fair value practitioners is to keep in mind that there should at least be consistency of assumptions when producing fair value adjustments for all those items requiring adjustments.

3.3. Should renewal business be considered in the fair value estimate and if so, how?
While future accounting guidance will include some discussion of what renewal guarantees are required for renewals to be included in fair value estimates, there undoubtedly will be areas of gray, such as how far a contractual provision regarding renewals has to go before it is considered a ***guarantee*** of renewal. For example, would a guarantee of a renewal at a price no more than the full policy limit (i.e. a riskless contract for the insurer) be considered a renewal guarantee?

3.4. How should judgement be accommodated in the development of fair value estimates?
All fair value methodologies have at least some judgmental elements within them. One of the objectives of fair value is to have the same liability held be two different entities have identical carrying values on each of the entities financial statements. The inclusion of judgement in the development of fair value estimates could result in situations in which different analysts are looking at similar liabilities but produce different results solely because of the judgmental elements.

4. Application of fair value methodologies – method-specific issues

Clearly from the pros and cons that accompany each of the methods discussed in sections C and D, no one method is appropriate in all situations. Each method has its strengths and weaknesses that may make it more or less appropriate as a technique for quantifying a liability's fair value. Rather than repeating the methods in sections C and D and identifying each method's implementation challenges, this section will describe implementation issues that are common across methods. A table summarizing the implementation issues associated with each method follows the descriptions.

4.1. Methods that rely on CAPM: as described in section D, the CAPM beta has been subject to criticism from both finance and actuarial sources. Finance theorists note that CAPM only recognizes nondiversifiable risk, assuming an efficient, friction-free market. However, insurance is not characterized by an efficient, friction-free market, which throws into question CAPM's applicability to insurance. Additionally, subsequent research has shown that more factors than just beta are needed to explain company stock returns. From the actuarial perspective, the concern is that estimates of underwriting betas have shown great volatility as well as the possibility of becoming negative.

4.2. Methods that rely on public data: not all companies' data is publicly available. This makes any method that relies on publicly available data subject to whatever distortions might exist from using a subset of all companies. Additionally, the data that is publicly available can contain distortions arising from systematic overstatement or understatement of liabilities by the entities providing the data. Lastly, there could be data compatibility issues arising from changes in the available data sets due to such things as mergers, insolvencies, divestitures, acquisitions, restructurings, etc. that alter the entities included in the data sets.

4.3. Methods that produce results only on a total company basis: if a method is used that produces results on an all-company basis but presentation requires that fair value results be displayed at a more detailed level, the methodology must be adapted to the presentation needs.

4.4. Time period sensitivity: the selection of the historical time period used as the basis for determining future parameters and assumptions could greatly influence the results.

4.5. Incorporates process risk: not all methods produce results that include a value for process risk.

4.6. Value additivity: not all methods produce results that are value additive.

4.7. Nature of the line of business: some methods are not well suited to the development of fair value estimates of liabilities arising from volatile short-tailed lines. All of the methods can be used for the development of fair value estimates of long-tailed lines' liabilities.

List of Considerations when Selecting an Estimation Method							
Method	Reliance on CAPM	Reliance on Public Data	Produce Results only on a Total Company Basis	Time Period Sensitivity	Incorporates Process Risk	Is Value Additive	Not Designed for Short Tail Volatile Lines
Undiscounted Value						X	
Present Value at a risk-free interest rate						X	
Present Value at a conservative interest rate						X	X
Entity-specific measurement						?	
Cost-accumulation measurement							
CAPM	X	X		X		X	X
Internal Rate of Return	X*			X		X	X
Single Period RAD	X*		X	X		X	X
Using Underwriting Results		X		X	X	X	
Based on Probability Distributions				X	X		
Based on Reinsurance		X**		X	X		
Direct Estimation		X	X	X		X	X
Distribution Transforms		X		X	X	X	
Naive Methods		X				X	?

* Can use other methods to develop the parameter input for the required return on equity.

** Public data is required when using public reinsurance quotes. Public data is not needed if the fair value estimates are derived from quotes made specifically for the entity that is developing the fair value estimate.

5. Presentation issues

The items presented here relate to the actual presentation of fair value results in a financial statement. These items are not "actuarial" in nature, but rather relate to the mechanics of financial statement presentation and disclosures required within the financial statement framework.

5.1. The selected method or methods may be appropriate for fresh-start valuations but not interim valuations. Fresh-start in this context refers to the accounting concept, not the tax one. The accounting concept of fresh-start involves "remeasuring an item using current information and assumptions" at each valuation date. (IASC Insurance Issues Paper, page A182.)

For example, suppose a company performs a full-scale actuarial review of reserves for a block of business twice a year. The company must publish financial statements quarterly, though. The liabilities booked after each full-scale review would be viewed as fresh-start valuations. However, for the financial statements produced between reviews, the company will need to have some other method of quantifying the proper liability value to record. The company can't just keep the same liability value from the previous financial statement. At a minimum, the company will need to adjust the recorded value to reflect payments made, unwinding of discount, and changes in the discount rate between the two statement dates. This process of updating the reported value without undergoing a full-scale analysis is an example of an interim valuation.

5.2. How should a restatement of historical exhibits to reflect historical fair value estimates be done? Any exhibits that show historical data would need to be restated to a fair value basis the first time fair value financial statements are produced. The question is how to do the restatement. Fair value should reflect conditions and market perceptions at the valuation date. It is difficult, if not impossible, to reconstruct these items after the fact, when what the outcomes of situations that were then uncertain are now known.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section G – Accounting Concepts

Introduction

This section discusses the proposed fair value adjustments in terms of the attributes demanded for sound accounting bases. We set out below the criteria (termed accounting precepts) that accountants and accounting standard setters judge accounting bases by, and consider who the users of financial statements are. We then consider each of the major fair value adjustments in terms of the accounting precepts. The fair value adjustment for the entity's own credit standing is discussed in section H.

Fair value accounting could be applied to any financial reporting; GAAP financial statements, statutory (regulatory) financial statements or even tax returns or internal management reports. While, in the U.S., GAAP financial reporting is determined by the FASB and the SEC, statutory financial statements will remain the responsibility of the NAIC. Even if fair value accounting were adopted for GAAP financial statements, a different non-GAAP basis might well be maintained for statutory financial statements.

Generally accepted requirements for 'good' accounting

The two relevant accounting pronouncements that discuss how to select the most appropriate accounting treatment from a range of alternatives are:

- FASB – Statement of Financial Accounting Concepts no. 2: Qualitative Characteristics of Accounting Information.
- IASC – Framework for the presentation of Financial Statements.

Fortunately, to a large extent the two documents agree as to what is desirable. The FASB document is longer and more discursive.

The IASC framework document defines the object of financial statements as:

> *"to provide information about the financial position, performance and changes in financial position of an enterprise that is useful to a wide range of users in making economic decisions."*

The desired traits of an accounting system are:

- Relevance
- Reliability
- Comparability and Consistency
- Neutrality
- Cost Benefit

Relevance. To be relevant information must be capable of making a difference to users' decisions. This is achieved either because the information can directly feed into a prediction of the future position of the enterprise, or because the information can be used to refine previous expectations. Untimely information generally has little relevance. The IASC framework details a separate characteristic of "understandability," stating it is an essential characteristic of financial statements that the information is readily understood by diligent users. This is implicit in the FASB concept of relevance, information which cannot be readily understood lacks the characteristic of being able to inform users' decision making. Also implicit in the two concept statements is the concept of transparency, i.e. that items in financial statements should be clearly disclosed so as to maximize their utility to financial statement users. (Neither the IASC nor FASB documents listed above mention transparency explicitly, although the IASC notes "substance over form," that is, following the economic substance rather than legal form as a basic requirement).

Reliability. Reliability depends on the representational faithfulness with which a reported item reflects the underlying economic resource, obligation or transactions. Reliability does not imply a need for certainty, and reporting the degree of uncertainty in an item may provide a better representation of the underlying economic reality than a single point estimate. In certain cases the measurements of the financial effects of items could be so uncertain that enterprises would not be allowed to recognize them in their financial statements (for instance, nonpurchased goodwill). Financial statements should be free from bias in their measurements. FASB, but not the IASC, notes verifiability as a characteristic that helps constrain bias in financial statements.

Comparability and Consistency. Financial statements should be comparable over time and between different enterprises in order to be able to ascertain trends and the relative position of different companies. Conformity to a uniform set of accounting standards helps achieve comparability and consistency.

Neutrality. Financial statements should be free from bias. However, the IASC framework notes that where an element of a financial statement is subject to uncertainty a degree of caution is needed in the exercise of judgment in making the required estimates.

Cost Benefit. The balance between cost and benefit is a constraint on "good" accounting paradigms rather than one of their qualities. If accounting information can only be generated at substantial cost, the relevance and utility of that information to users needs to be established before it is sensible to adopt accounting standards that demand such information.

Fundamental Assumptions

The IASC framework notes two fundamental assumptions for the preparation of financial statements. These are:

- ***The Accruals basis***: Transaction are recognized when they occur, not when cash changes hands, and reported in the financial period to which they relate.

- ***The Going Concern basis***: Financial statements are prepared on the basis that the enterprise will continue in business for the foreseeable future. If there is the likelihood or intention to substantially curtail business or to cease to trade, financial statements may need to reflect this in their choice of accounting policies, and the circumstances are to be disclosed.

Accounting paradigms

There are two types of modern accounting paradigm.

There is the ***deferral-matching*** approach, such as in traditional property casualty accounting. This approach can be characterized as income statement focused. They aim to match revenue and expenses of a period in the income statement of that period, and "park" surplus contractual income flows (future income) and surplus costs (such as deferred acquisition costs) in the balance sheet so they can be reflected in a subsequent periods' income flows.

The alternative is the ***asset-liability*** approach. These models are balance sheet focused. Their aim is to accurately reflect the assets and obligations of a company at periodic intervals. The changes in the values of assets and obligations become the profit (or loss) for that period. A fair value accounting approach for the assets and liabilities of insurance enterprises is one potentially available asset-liability paradigm.

The IASC paper essentially analyses three alternative methods of accounting for insurance: the current deferral-matching model, full fair value accounting, and an alternative asset-liability model.

Who uses financial accounting, what are their needs, and on what do they focus

Shareholders, analysts and potential capital providers

Shareholders and potential capital providers fall into two classes, the professional, often institutional, investor and the individual investor. Both may be interested in the long-term earnings potential of the stock, or the potential for short-term capital gains from holding the stock. Both groups will be interested in earnings trends, the adequacy of reserves for future payments and the value and quality of assets held. Sophisticated users should be able to unravel almost any accounting treatment given sufficient disclosure, (although whether they will in practice be attracted to doing this is questionable). For unsophisticated users it is highly desirable that trends in current earnings can be distinguished from fluctuations arising from volatile shifts in fair value measurements. In addition, they may find it useful to have clear indications of balance sheet risk. Sophisticated users are also likely to welcome user-friendly presentation, particularly in the income statement, and clear indications of balance sheet risk.

Policyholders, potential policyholders, brokers and rating agencies

Personal and some small commercial policyholders are unlikely to resort to examining insurers financial statements before purchasing insurance. If they use an independent broker for their purchase, the broker is more likely to rely on rating agencies' assessments than to carry out their own assessment of insurers.

Most prospective commercial insureds and reinsureds and their brokers are interested in the solidity of (re)insurers with whom they place business. Essentially they need to evaluate the risk of the (re)insurer being unable to pay claims in full once they become due. While income statement information is not irrelevant, their basic focus is on the balance sheet strengths and weaknesses of the company.

Existing commercial policyholders and in particular policyholders with outstanding claims against insurers/reinsurers of doubtful solvency, require that financial statements provide them with sufficient information to evaluate the credit risk they face from their existing policies' receivables, so that they may plan and act accordingly.

Rating agencies have similar aims as commercial insureds and reinsurers. Their basic focus is on balance sheet solidity. They, like insurance sector analysts are sophisticated users of financial information, and have access to more detailed financial information than that presented in the financial statements.

Bankers and Other Creditors

Bond issues and bank loans are most likely to be the obligation of the holding company of insurance groups, not the individual insurance entities underneath the holding company. The bond holders and bankers behind this debt will be interested in the ability of insurance groups to service borrowings and repay loans, this is a function of both balance sheet strength and the future profitability of the company. In addition both these creditor groups may be interested in ascertaining that covenants are satisfied.

Regulators

Regulators have, at least in the US, two perspectives on insurance companies. First, they are interested in the solidity of insurance companies and in minimizing any call on guarantee funds. Second, they may wish to use the financial statements as a resource in the regulation of prices. Regulatory analysis in both these areas might be made more difficult if reported profit measures are volatile. Well understood and accepted measures of shareholder equity would also be advantageous. Regulators have access to other financial information. Indeed, in the US, statutory financial reports will be their primary source for the financial review of an insurance company's operation.

Outside the US, regulators make more use of a company's general purpose financial statements, and generally desire a single accounting paradigm for general purpose and regulatory financial reports.

Employees

Employees will be concerned primarily with two questions: how secure is the company? and how well is it doing? Most employees will be unsophisticated users of financial statements.

<u>Discussion of fair-value valuation bases in the context of accounting precepts.</u>

Fair value adjustment – marking investments to market.

The principal actuarial issue associated with marking of investments to market is balance sheet consistency. If investments are marked to market, then their value will fluctuate with various financial variables, such as interest rates. If the same variables also impact the economic value of the liabilities, but not the stated value per accounting rules, then reported income and equity will be distorted. These reported income and equity values, and especially the reported changes in those values, will not be relevant and will not be representationally faithful.

If insurance company investments are recorded at fair value, then reporting insurance liabilities at fair value will create consistent balance sheet accounting, and will improve relevance and representational faithfulness of reported income and equity.

There are alternatives to fair value accounting for liabilities that react to some, if not all, of the same variables impacting the investment market value. These alternatives may produce more relevant financial reports than the current status quo for U.S. GAAP (where most liabilities are undiscounted but many assets are at market). They may also be easier to implement that full reflection of fair value for liabilities. The risk is that they may cause an unacceptable level of inconsistency relative to the assets, for those financial variables that would impact market values but not the alternative standard liability values.

Fair value adjustment – discounting
(as applied to loss and expenses reserves, reinsurers' share of loss reserves, unearned premium reserves and possibly debtor balances and deferred taxation.)

Currently, most p/c reserves are carried at an ***undiscounted*** value. This current use of undiscounted reserves for loss reserves has the following advantages and disadvantages.

Advantages

- It is easy to understand
- It locks in a margin that cannot be distributed to shareholders. (A plus in the eyes of regulators and policyholders)

Disadvantages

- It is typically an unreliable measure of the economic value of liabilities. Further, the degree of distortion varies between different enterprises depending on their mix of business and growth history. As a result, return on equity comparisons are distorted both within the insurance sector and with other industries. In particular, insurance company equity is understated in most cases compared to values for other industries. This understatement of insurance company equity leads to an overstatement in returns on equity.
- It results in different valuation bases for assets and liabilities, which can result in spurious earnings volatility when interest rates change even when the underlying cash flows are broadly matched.
- It distorts profit recognition.
- Booking undiscounted reserves may provide grounds for accounting arbitrage.

Fair value proponents, and others in favor of moving to a ***discounted*** basis for insurance liabilities, would argue that moving to a discounted basis for loss reserves, etc., removes or at least substantially reduces:

- The inconsistency between the valuation basis of assets and liabilities, to the extent assets are either at market or at some version of cost (which is effectively an historic market value).

- The inconsistency between enterprises writing different classes of business where the economic value of two reserves shown at the same amount may be substantially different.
- The conservative bias that may be implicit in undiscounted liability values.

They would argue that the profits reported on a discounted basis would be a better (more relevant) reflection of an enterprise's earnings for a period. The use of a fair value liability valuation (in conjunction with holding assets as market) will put assets and liabilities on a consistent footing, so that changes in the values of assets and changes in the discounted value of liabilities broadly mirror each other when interest rates change, so long as liabilities and assets are matched. This will eliminate that part of the interest rate volatility that does not reflect economic change for the insurance enterprise. Further, fair value proponents would maintain that the balance sheet values calculated on a discounted basis better discern between different enterprises; that is they are more relevant, and do not contain conservative biases; that is they are neutral.

Fair value proponents would also argue that well thought out presentation in the income statement matching of investment return and the unwinding of the discount could do much to mitigate the potential confusion that may be suffered by some users as a result of moving to a discounted basis for loss reserves.

Others who oppose the introduction of discounted amounts would argue that liability values currently reported by insurers reflect two offsetting biases, i.e., lack of provision for future investment income and optimistic evaluation of ultimate settlement values (resulting in insurance liabilities that they believe are already implicitly discounted). The introduction of explicit discounting would remove one of the two biases. However, valuing loss reserves at discounted values without addressing the second bias would probably be a disservice to all users as it would overstate available capital and overstate profitability.

Further such observers might argue that if fair values are assessed by direct comparison to exit prices available in the reinsurance market, there is a danger that values substantially different from the net present value of the cost to the enterprise of running off liabilities may be recorded. Substantial overvaluations are possible when there is a hard reinsurance market. Substantial undervaluations are possible when there is a soft reinsurance market, precisely the time at which such valuations cause regulators most concern.

The use of discounted liabilities will not necessarily result in more or less reliable estimates than the undiscounted ones. Discounting techniques are well understood and generally introduce little additional subjectivity into the liability valuation process. When the uncertainties are concentrated in the tail, discounting of the reserves may even reduce the uncertainty in the estimated liability value. In this task force's opinion, fair value accounting in practice may not significantly alter the inconsistency between different company's accounts due to variations in reserve strength.

Essentially similar arguments apply to the introduction of discounting for the estimates of other insurers' liabilities or assets

Fair value adjustment – risk margins
(as applied to loss and expenses reserves, reinsurers' share of loss reserves and unearned premium reserves.)

Fair value proponents would argue that discounting in conjunction with adding risk margins to liabilities provides the best basis for profit recognition. The profit on the book of business will emerge as the associated risk expires.

This approach has the drawback that it is a difficult concept to grasp and may confuse amateur (and some professional) users of accounts. Clear disclosure of the risk adjustment may help such users.

The lack of market depth in the exchange of insurance liabilities between enterprises makes a direct market assessment of the price for the risk margin impossible in most instances. Risk adjustments derived from methods that use industry-wide data to derive industry level risk adjustments may not succeed in producing financial information that can be used to distinguish between one insurance enterprise and its peers. In addition market-based information will be impossible to obtain in countries that do not have significant stock markets, or that have integrated financial service industries where the major insurance carriers also have banking and securities interests within one quoted vehicle.

Other enterprise-specific risk measures can to a greater or lesser extent be criticized as requiring significant subjective input. Proponents of such methods would argue such judgment calls are inherent in arriving at other accounting measures such as the bad debt adjustment to trade receivables in manufacturers' balance sheets.

This is an area where standard setters may well be faced with determining a trade off between reliable (less subjective) and relevant measures.

If there is a wide range of acceptable methods for calculating fair value adjustment this may well lead to a greater spread of the range of acceptable "values" for the various elements of financial statements. Accounting/actuarial guidance is likely in practice to increase the consistency of the calculation of the risk margin.

The introduction of subjective elements into fair value assessments also means that there is additional scope for managing (or manipulating) financial results. Methods that reduce the scope for subjectivity in the assessment, such as an IRR model using regulatory capital, curtail the scope for inconsistency between different insurance enterprises (but, possibly, at the expenses of relevance, see above). More company specific methods may result in greater scope for inconsistency (the scope might well in practice be reduced by accounting or actuarial guidance).

The task force suspects however that the increase in inconsistency due to differences in the basis on which fair values are calculated are likely to be of second order compared to differences in the strength of company's loss reserves.

Opponents of risk margins would argue that a risk margin for insurance liabilities cannot be reliably determined, so that (per FASBs Concepts Statement No. 7, paragraph 62) discounted values with no risk adjustment should be used. Others would argue that undiscounted values would be preferable to discounted values without risk adjustments, which they would contend, could grossly understate a company's liabilities.

Fair value adjustment – To reserves and creditors to reflect a company's own credit standing.

This is the most contentious of the fair value adjustments, and is separately discussed in section H.

Taxation

The extent of the link between taxes and the financial statements of enterprises varies between different countries. Where the calculation of taxable profits is substantially based on the profit disclosed in the enterprise's general purpose (i.e., GAAP) financial statements, it is certainly possible that at least some companies may suffer a greater burden of taxation. It is possible this may be mitigated to some extent by the recognition for tax purposes of some allowance (i.e., risk margin) for the uncertainty in estimated claim liabilities. In the U.S., the explicit recognition of risk margins may cause them to be removed from allowed claim liability deduction, thereby increasing federal income taxes unless the margins are allowed by the IRS as a part of the liabilities' economic value. If the reserves are currently reported at expected value, the risk margins would have no impact on taxes (if the margins are accounted for as an asset) but would restrict the disposable income.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section H – Credit Standing and Fair Value Liabilities

A highly controversial proposed adjustment to estimated cash flows in the determination of fair value liabilities is the impact of the entity's (or obligor's) own credit standing. Under some proposals, the weaker the obligor's financial situation, the lower the fair value of their liabilities would be. This adjustment would recognize that a financially weak company would be less likely to satisfy its obligations in full than a financially strong company.

This issue may not be material for most insurers, as it is very difficult for an insurer to be both viable and of questionable financial health. Companies viewed to be strong financially have historically experienced very small rates of default.[56] Therefore, the concern and controversy surrounding this issue is focused largely on its impact on troubled companies.

This section of the white paper presents the arguments for each side of the issue, without stating an overall preference. It also discusses the issues associated with estimating, implementing and presenting liabilities that reflect the obligor's credit standing.

This section is organized as follows:

- Arguments ***for*** reflecting credit standing in fair valuing liabilities.
- Arguments ***against*** reflecting credit standing in fair valuing liabilities.
- Methods for estimating this effect.
- Presentation issues.
- Implementation issues.

Arguments for reflecting credit standing in fair valuing liabilities.

- Credit risk is reflected in the fair value of assets, and the assets and liabilities should be valued consistently.
- The public debt of a company has a market value, and that market value reflects the debtor's credit standing. Hence, requiring a company to report their publicly issued debt (a liability for them) at market value leads to requiring them to reflect their own credit standing when valuing a liability. The alternative, not requiring a company to report such debt at market value, would allow a company to manipulate its earnings by buying back existing debt or issuing new debt.
- If public debt is to be held at a fair value that reflects credit standing, then all liabilities should be reported at a fair value that reflects credit standing. This is the argument FASB made in their Concepts Statement Number 7, paragraph 85.
- Parties owed money by a company of questionable solvency will frequently settle for less than the stated amount of the obligation, due to the risk of possibly getting much less if that company (i.e., the obligor) goes insolvent. In other words, reflecting an entity's own

[56] One year default rates for debt rated A or above (by Moody's) were less than 0.1%, for 1983-1999. Ten-year default rates for the same rating category were less than 4%, for 1920-1999. Source: January 2000 report by Moody's on Corporate Bond defaults from 1920-1999.

credit standing in valuing its liabilities reflects the true market cost to settle those liabilities.

- The obligor's credit standing is easily measurable, at least in those jurisdictions where established rating agencies exist.
- Due to limited liability, the owners' interest (e.g., as reflected in share price) of a company can never go below zero. Thus, the fair value of its equity is always greater than or equal to zero. If the fair value of the equity is greater than or equal to zero, and the fair value of the assets is less than the contractual "full value" liabilities, then the fair value of the liabilities must be less than this "full value."

Arguments against reflecting credit standing in fair valuing liabilities.

- There is no active market for such liabilities; hence there is no reliable way of measuring this adjustment for credit standing.
- Users of financial statements could be misled as to the financial strength of weak companies.
- A liability valuation that reflects the liability holder's credit standing would not be relevant to a potential "buyer" of the liability. In the insurance situation, and possibly other situations, the buyer would not be able to enforce the same credit standing discount on the obligee. The obligee would view the prior liability holder's credit standing as totally irrelevant. Hence, the buyer would also view the credit standing of the liability seller as irrelevant to the liability's market value.
- An obligor's financial statements that included a reduction in the fair value of its liabilities due to the obligor's credit standing would not be relevant to creditors.
- An insurance company's principal product is its promise to pay. In return for cash up-front, an insurance company sells a promise to pay in the event of a specified contingency. If an insurer attempts to pay less than the full initial promise, due to its weakened credit standing, it is in effect abandoning its franchise. In fact, a troubled company that is trying to remain a going concern will do all it can to pay the full amount, in an attempt to retain its franchise. As such, reflection of credit standing in the estimation of fair value liabilities is counter to going-concern accounting, and is relevant only to liquidation accounting for a runoff business. (The party trying to collect from a troubled company is also arguably negotiating under duress. As such, any settlement amount they would arrive at would not meet the definition of "fair value.")
- If credit standing is reflected in liability valuation, then favorable business results could cause a drop in earnings, due to an improved credit standing increasing the fair value of liabilities. Likewise, unfavorable results that lead to a drop in credit standing could result in earnings improvement. This is counterintuitive and noninformative.
- It does not make sense to reflect credit standing in the value of liabilities without also reflecting the impact of credit standing on intangibles. A company with a worsening credit standing may see the fair value of its liabilities decrease, but it would also see the fair value of various intangibles, such as franchise value, decrease. In fact, the existence

of the intangible franchise value helps keep insurers from increasing their operational risk in order to increase shareholder value at the expense of policyholders. Therefore, while the fair value of a company's liabilities may be decreasing as credit standing decreases, it is offset by an item not to be reflected in the fair value accounting standards as currently proposed by the FASB and IASC. If intangibles are not to be estimated nor reflected in a fair value standard, then the impact of credit standing on the liabilities should not be reflected.

- Credit standing is (usually) an attribute of the corporate whole, not the individual business segments. Hence, business segment reporting could be complicated drastically by this approach, as the segment results would not add to the corporate whole without an overall credit standing adjustment.
- To the extent that the credit standing adjustment is based on the obligor's judgment, a potential moral and ethical dilemma exists. Management may be forced to state the probability that it won't pay its obligations at the same time that it may be professing before customers, partners, capital providers, etc. its integrity, financial soundness and full intent to meet all obligations.
- If an entity's own credit standing is reflected in valuing their liabilities, and the valuation considers the reduced amounts their policyholders may be willing to accept as claim settlement, some companies may be motivated to employ unreasonably optimistic assumptions in setting their reserve levels. Troubled companies may be incented to anticipate that claim settlements will be resolved on extremely favorable terms and hence record an inappropriate reserve.

Methods for estimating the impact of credit standing on liabilities, if included in the fair value definition.

Our task force was able to envision several methods that might be used to estimate this credit risk adjustment. Four such methods are listed here. It is important to note that, to our knowledge, none of these methods have actually been used to estimate the fair value of liability default for property-liability insurers in any practical setting. The first three methods are discussed in more detail in the appendix, including examples.

Method 1 - Implied Option Value

The reflection of credit standing in the valuation of fair value liabilities (i.e., the "credit risk adjustment") involves estimating the expected fair value of liability default. In the finance literature, the default value has been shown as equivalent to a put option on the insurer's assets.[57]

[57] Cummins, J. David, 1988, Risk-Based Premiums for Insurance Guaranty Funds, Journal of Finance, September, 43: 823-838. Also,
Doherty, Neil A. and James R. Garven, 1986, Price Regulation in Property-Liability Insurance: A Contingent-Claims Approach, Journal of Finance, December, 41: 1031-1050. Also,
Derrig, Richard A., 1989, Solvency Levels and Risk Loadings Appropriate for Fully Guaranteed Property-Liability

Thus, the theory underlying the credit risk adjustment (in the insurance context) is that the fair value of owners' equity is increased by the value of the option implicitly given to the equity owners by the policyholders. If the liabilities are measured without the credit risk adjustment, then the fair value of the owners' equity is understated.

The implied option value can be determined by the method of Ronn and Verma[58], which is used in the Allen, Cummins and Phillips analysis.[59] Under this method, the market value of the firm's assets is first estimated. Then the implied volatility of the firm's market value is estimated from the Black-Scholes formula for the value of the equity owners' call option.[60] Other inputs required for this estimation are the undiscounted liability value, the average time until payment of the liabilities and the risk-free interest rate.

Once the above inputs are obtained, the default value is determined by applying the Black-Scholes option model with a set time to expiration and an exercise price equal to the expected liability value at the end of the same time horizon. The call option is valued relative to the asset market value. The Appendix provides an example of the calculation.

Advantages

- For publicly traded insurers, this approach can provide results using an insurer's own data.
- The method is relatively straightforward in terms of the complexity of the calculation.
- The method has been used to measure default risk for both insurance firms and banks. It is well known in the finance literature.

Disadvantages

- This method can only be done for publicly traded companies.
- It is difficult to carve out the property/casualty pieces of firms that have non-property/casualty business segments.
- The method is sensitive to variations in input values.
- The method relies on accounting value of liabilities. This presents problems with measuring reserve adequacy.
- It ignores side guarantees or implicit guarantees, such as that from a majority owner with a reputation to uphold. Such an entity cannot afford to walk away without losing brand-name value. It also ignores the side guarantee arising from an insurance guaranty fund.

Insurance Contracts: A Financial View, Financial Models of Insurance Solvency, J. D. Cummins and R. A. Derrig eds., Kluwer Academic Publishers, Boston, 303-354. Also,
Butsic, Robert P., 1994, "Solvency Measurement for Property-Liability Risk-Based Capital Applications", Journal of Risk and Insurance, 61: 656-690.

[58] Ronn, Ehun I., and Avinash K. Verma, 1986, Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model, Journal of Finance, 41(4): 871-895.

[59] Allen, Franklin, J. David Cummins and Richard D. Phillips, 1998, "Financial Pricing of Insurance in a Multiple Line Insurance Company", Journal of Risk and Insurance, 1998, volume 65, pp. 597-636.

[60] Black, Fischer and Myron Scholes, 1973, The pricing of Options and Corporate Liabilities, Journal of Political Economy, May-June, 81: 637-659.

- It may ignore the relative credit-worthiness for different lines or entities within the corporate total, if they have separate publicly traded securities.

Method 2 - Stochastic modeling using Dynamic Financial Analysis (DFA)

Stochastic modeling is frequently used in Dynamic Financial Analysis to model insurance company operations. The process typically involves modeling assets, liabilities and future income from the runoff of reserves as well as new business. Key variables driving outcomes are modeled using probability distributions.[61] In addition to projections of future cash flows, stochastic DFA models can produce Statutory and GAAP balance sheets and income statements.

DFA models attempt to incorporate the dynamics of the insurance business by including interactions between the different variables. Some DFA models also attempt to model the underwriting cycle.

Among the outputs of stochastic DFA models are probability distributions of future surplus. They can be used to compute the expected policyholder deficit (the expected cost of default), or the average amount of unpaid liabilities, should the company experience insolvency in the future. Insolvency would be deemed to have occurred whenever the company's surplus dropped below a pre-specified level.

Advantages

- The method is insurer-specific.
- The method can be applied to all insurers.
- A comprehensive DFA model can better incorporate important company-specific risk factors than the other methods.
- Many companies currently use these models to make strategic business decisions. A great deal of research effort has recently been devoted to their development.

Disadvantages

- Good DFA models tend to be complex and are therefore labor-intensive and expensive. (However, if an insurer already has such a model, adapting it to estimate credit risk may require little additional cost.)
- DFA models are designed to work off of data. They may not reflect risks that are not in the historical data.
- Not all insurers currently have these models, since their management has determined that they are not worth the cost. Insurers would need the models to be tailored to the unique features of their business.
- There is presently not enough expertise available to construct a suitable DFA model for each insurer.
- The models may not produce comparable results for similar companies, due to different model structures and parameter assumptions.
- The ability of these models to reliably estimate insolvency probabilities is not universally accepted. Many believe that these models are stronger at estimating the normal variation

---

[61] This is a feature of stochastic DFA models, but not necessarily all DFA models.

resulting from the current processes, and not the shocks and paradigm shifts that may be more likely to be the cause of an insolvency. Therefore, they may not be reliable when applied to the stronger companies (although these companies are not expected to have a material credit-standing adjustment).

- It may be impractical to model insolvency for large, multinational or multi-industry conglomerates.
- Business and legal problems may exist for companies estimating their own probability of reneging on their obligations, either directly or through a DFA model estimate.

Method 3 - Incorporate historic default histories by credit rating from public rating agencies.

This method would use publicly available historic default rates by credit rating, based on the entity's current credit rating from A. M. Best, S&P, Moody's or some other public rating service. At least one of these rating services (Moody's) publishes historic default rates by credit rating, for a one year and multiple-year horizon, by year and averaged over several decades. These default rates would allow determination of the expected default rate — some other method would have to be used to determine the risk premium associated with this expected value.

Advantages

- Simple to use and explain, when using the expected cost of default from the public data.
- Requires little direct analytical cost to the insurer.
- Avoids an entity having to estimate its own probability of reneging on promises.

Disadvantages

- Ambiguity would exist if the various public ratings are not consistent. For example, it is common for the ratings from Moody's and S&P to differ. This would add judgment to the process and potential manipulation.
- Not all companies are rated.
- A single rating may exist for the enterprise (such as a group rating), that may not be appropriate for a particular group member or a line of business.
- Would require default history for a given rating. These may not be available from some rating agencies.
- Requires ratings to be consistently applied over time. This may not be the case, as rating methodologies change over time.
- Ratings may exist for debt, but not for all other liabilities. This problem could be compounded by the existence of guaranty funds, particularly where those guarantees vary by state and line.

Method 4 - Utilize credit risk-based spreads observable in public debt.

This method would utilize observed interest rate spreads on public debt to quantify the credit risk adjustment. Public debt has no amount risk, other than default risk, and no timing risk (absent call provisions). Hence, it can be used to isolate the market's pricing of credit risk. The discount that the market places on a dollar owed at time X, given a credit rating of Y, compared to the same market value for a dollar owed at time X by the U.S. government, quantifies the credit risk adjustment for a time horizon of X, rating of Y.

Ideally, this would be done based on the market value for each company's publicly held, noncallable debt. If not available, then public debt of companies with a similar credit standing (as measured by a public rating agency) could be used instead.

It may also be possible to use the developing market for credit derivatives rather than public debt in applying this approach.

Advantages

- Relatively simple to use and explain.
- Requires little direct analytical cost to the insurer.
- Avoids an entity having to estimate its own probability of reneging on promises.
- Consistent with credit risk adjustment for public debt issued by the same entity.
- Relies heavily on market-based values rather than internal estimates.

Disadvantages

- Requires information on a range of public debt instruments that may not exist for all companies. The entity may not have any actively traded public debt, or may not have a broad enough range of noncallable public debt to handle all the time horizons of interest.
- Where reliance is made on other entities' public debt with similar credit standings, it requires a determination of whether or when another entity has a similar credit standing. This adds additional judgment and estimation to the method.
- Debt holders credit risk is not perfectly aligned with policyholder credit risk. Due to the different priorities of creditors in a bankruptcy or insolvency proceeding, the amount recoverable under a bankruptcy could be drastically different for policyholders as opposed to debt holders. In addition, since debt is frequently at the holding company level, it is possible that the bankruptcy administrator could arrange for a buyer to take over the insurance operation such that the policyholders would be made "whole", at the expense of the debt holders.
- Does not allow for guaranty funds or other side guarantees not applicable to public debt. These guaranty funds and side guarantees can also vary by state and line, further distancing the public debt information from the task at hand.
- The public debt may only exist for the enterprise (e.g., parent or holding company), which may include many other businesses and operations besides the insurance operation. The net credit risk may actually vary drastically by operation, so that the enterprise's public debt credit risk is not indicative of the insurance operation credit risk.
- To the extent that the observed debt is callable, this could distort the application of observable spreads to liability credit standing adjustments.
- Observed spreads versus U.S. Treasuries could include factors other than credit risk, such as relative liquidity.

Presentation issues.

The following are a few presentation issues surrounding the reflection of credit standing in the fair value of liabilities, assuming that such a reflection is made.

- ***Historical loss development*** - Should historical loss development include the impact of changing credit ratings (of the liability holder)? Choices are to include this impact, to exclude this impact, or include this impact but separately disclose this impact.

- ***Current balance sheet impact*** - The task force generally agreed that the current impact of credit standing reflection on the balance sheet should be disclosed, so as to provide useful information for those interested in the total legal obligations of the entity.
- ***Impact on income*** - Should the impact of credit standing reflection be separately disclosed when reporting period earnings?
- ***Impact on segment results*** - Most financial statements include various types of "segment" disclosures, i.e., disclosures about certain business or operating segments of the business. Current U.S. statutory reporting also includes many disclosures by product or line-of-business. Where a corporation's debt is held principally at the holding company corporate level, and not at the segment or operating level, it many not be appropriate to reflect credit standing adjustments in business or operating segment results. In such a case, credit standing adjustments would be reported only at the total corporate level, as an overall adjustment to the business segment "pieces." Alternatively, credit standing could be incorporated at the business-segment level, at the cost of potentially misstating the earnings or value of the business segment.

  If reported at the business-segment level, credit standing adjustments could distort reported business-segment results in another way. Consider the case where most debt is at the holding company level, the total corporate credit standing is weak, and the principal cause is a single business unit. If credit standing is reported at a detail level, operating earnings of the stronger business units would be impacted by the results of the unrelated, poorly performing unit. Worsening results in that poorly performing unit could lead to improved earnings (due to reduction in liability valuations) for the stronger units, while improving results for the poorly performing unit could cause lower earnings for the stronger units.

Implementation issues. The following are some possible implementation issues associated with reflection of credit standing in fair value estimates.

- ***Multiple credit standings.*** - It is possible for the different entities in a corporate whole to have different credit standings. For example, it is conceivable that the flagship of a quota share pool may be weaker than one of the quota share pool members. In such a case, it may be difficult to quantify all the differences, especially if all the publicly available data regarding credit standing is applicable only for the pool flagship.
- ***Incorporating credit standing adjustments when multiple risk adjustment methods are used.*** - Section D discussed several different methods for estimating the fair value risk adjustment. It is possible a single company would find itself using different methods for different lines. It may be difficult to incorporate the chosen credit standing adjustment consistently into the results of these various methods.

- ***Consistent treatment where offsets exist***. - Some liabilities have corresponding offsets, recorded either as assets, contraliabilities, or even as other liabilities. Examples include accrued retro premiums for retrospectively rated business, deductible recoverables, and contingent commissions. If a liability is valued in a manner that reflects the obligor's credit standing, then the valuation of offsets for that liability should also be impacted in a consistent manner. This may not be a simple task, and may materially complicate the estimation process for both the direct liability and the offsets.
- ***Guaranty fund reflection***. - The credit standing adjustment of a liability could be materially impacted by any guaranty fund (or similar) protection. The rationale is that the party owed money (e.g., a claimant) may be unwilling to consider lowering their cash settlement demands despite the financial weakness of the obligor, to the extent that there is backup protection provided by a guaranty fund. Guaranty funds do not exist for all lines nor in all states. They typically provide less than full protection (e.g., many funds cap the benefits, and may pay claims only after significant delays). As such, proper reflection of guaranty fund impacts may be very difficult, especially for a writer of multiple products in multiple states.
- ***Management dilemmas*** - It may be difficult for management to value its liabilities reflecting less than full contractual obligations, at the same time it is making assurances and promises to consumers and creditors, especially when the impact of the credit standing is significant.
- ***Auditor dilemmas*** - Whoever audits a company reporting fair value liabilities lowered for credit standing impacts may find itself in the same position as a rating agency. That is, it may be forced to quantify the likelihood of client solvency when auditing their financial statements. This may be outside their normal expertise, and could open up additional areas of auditor liability.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section I – Professional Readiness

Previous sections of this white paper have discussed what fair valuing means, what methods can be used to accomplish it, and what theoretical and practical issues must be dealt with in order to implement the fair valuing of insurance liabilities. This section discusses what the actuarial profession needs to do to prepare for its role in this process. Evaluating what casualty actuaries need to do to prepare for fair valuing insurance liabilities requires addressing the following four issues:

- Do actuaries currently have a theoretical understanding of fair value concepts adequate to estimate liabilities under a fair value standard?
- Are models currently available that can be used by actuaries to estimate fair value liabilities?
- Are actuaries prepared to implement these models and make these estimates in practice?
- What steps can the profession take to aid individual actuaries in implementing effective processes for fair valuation of insurance liabilities for their companies or their clients?

Note that professional readiness for this task should be evaluated relative to a hypothetical implementation date sometime in the future. Fair valuing insurance liabilities is not currently required of insurers in the United States, and we assume that initiation of such a requirement would be accompanied by a reasonable implementation period.

Adequate theoretical understanding and appropriate models

The analysis done by the task force and presented in the preceding sections demonstrates that actuaries have the theoretical understanding needed to implement fair valuing of insurance liabilities. We have identified a number of models that are available and appropriate for actuaries to use in estimating fair value liabilities. No issues have been identified that are not susceptible of actuarial estimation.

Ability to make estimates in practice

As noted above, fair valuing insurance liabilities is not a current requirement for most insurers in the United States. Therefore, actuaries generally have not established the systems and procedures that would be required to efficiently support fair valuation of liabilities for the financial reporting process. However, casualty actuaries performing insurance pricing and corporate financial functions have used many of the fair value models that have been identified in prior sections of this white paper, and the task force believes that this precedent demonstrates that actuaries can estimate fair value liabilities in practice.

The task force has identified a number of issues concerning fair value that require clarification prior to implementation. The task force presumes that many of these issues will be clarified later in the accounting standards development process. The task force also presumes that a reasonable period will be provided for implementation of any new accounting standard requiring fair valuing insurance liabilities. Given those assumptions, the task force believes that actuaries will be able to develop and use models that provide efficient and effective estimates of the fair value of insurance liabilities in accordance with those new accounting standards.

Steps the profession can take

The task force believes that there are a number of steps that can and should be taken by the actuarial profession to aid individual practitioners if fair value accounting for insurance liabilities is adopted for U.S. GAAP or statutory accounting. Depending on the course of future accounting standards developments, the same may be true if the IASC adopts fair value accounting for insurance liabilities.

1. You hold in your hands the first step, a white paper that discusses fair valuation of insurance liabilities for general or property/casualty insurers. The task force hopes this document will aid accounting standards setters in developing higher quality standards for insurers. The task force also hopes this document will be a starting point for casualty actuaries seeking both to better understand the issues underlying fair value accounting and to plan what methods to use in fair valuing insurer liabilities for their own companies or clients.

2. The actuarial profession should continue its active participation in the ongoing discussions of fair value accounting for insurers. As is evident from the prior sections of this white paper, fair value accounting is a complex issue, and actuaries should continue to provide active assistance to accounting standards setters in order to insure that the adopted standards are of high quality and are practical to implement.

3. The profession should seize any opportunities to broaden the numbers of actuaries engaged in the discussion of fair value accounting. CAS meetings and the Casualty Loss Reserve Seminar (CLRS) are the most obvious opportunities to discuss these concepts with more casualty actuaries. Publication of this white paper in the CAS *Forum*, on the CAS web site, and in other appropriate public forums should also be encouraged.

4. Once an accounting standard setting organization adopts fair valuing for insurance liabilities, a practice note designed to highlight the issues that practicing actuaries may wish to consider in implementing that standard should be produced as soon as

possible. Practice notes are designed to provide helpful information quickly, so they do not go through the due process required of a new Actuarial Standard of Practice (ASOP). Accordingly, neither are they authoritative for actuaries. In addition to being published, any such practice note should be presented at the CLRS and at CAS meetings.

5. Finally, the task force believes that issues will arise during implementation that have not been anticipated in advance. Initially these should be handled through updates to the practice note. Once some experience has been accumulated, there may be need for consideration of a new or revised ASOP. The task force has not identified any need for a new or revised ASOP at this time and believes it is better to defer developing any such standard until actual practice under a fair value accounting standard has had a chance to develop. Premature development of an ASOP may mean that unanticipated but important issues are not addressed in the ASOP. Also, an ASOP developed too soon may tend to impede the development of good practice by requiring more justification for estimation methods not yet contemplated during the drafting of the ASOP.

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities

Section J – Summary and observations

This white paper has discussed many of the major issues involved in fair value accounting as applied to insurance liabilities. While the focus has been on property/casualty insurance liabilities, many of the issues are also applicable to other insurance liabilities.

In brief, some are the major findings of this paper include:

- ***New task*** - Generally, fair value accounting rules have not yet been applied to property/casualty insurance liabilities. Therefore, implementation of fair value accounting would likely result in unforeseen consequences and a learning curve for those charged with implementing the new rules. (One way to address this issue may be to field test a fair value accounting system before full implementation, possibly via footnote disclosure.)
- ***More work*** - Implementation of fair value accounting for these liabilities would be an increase in workload for those setting the liabilities. New systems and procedures would have to be set up, and additional estimation variables would have to be monitored.
- ***More assumptions*** - Fair value accounting would increase the number of subjective assumptions required for most property/casualty reserving. The impact of these additional assumptions, however, may still be of second order importance when compared to the variability across companies in the (undiscounted, pre-risk adjustment) expected loss estimates.
- ***Multiple methods*** - A critical component in fair value estimation is estimation of the risk margin, or risk adjustment. There are several methods that can be used to estimate these risk adjustments. Each method has advantages and disadvantages, and, depending on the variation in liabilities to be estimated, the use of multiple methods may be necessary.
- ***Can be done*** - No issues have been identified that are not susceptible of actuarial estimation.
- ***Not without concerns*** - As mentioned previously, problems would undoubtedly occur during any initial implementation of fair value accounting.
- ***Evolutionary process*** - Familiarity, expertise and available methods for estimation of fair value liabilities should grow over time, once fair value accounting for insurance liabilities is implemented. Many of the initial estimation problems should diminish over time.
- ***Presentation and Implementation issues, in addition to estimation issues*** - There are issues besides strict estimation issues that actuaries (and accountants) will have to deal with. These include questions as to how historic loss development should be presented in a fair value paradigm, and whether the lack of "value additivity" is an advantage or disadvantage of an estimation method.
- ***Alternatives exist*** - There are other accounting paradigms besides the fair value paradigm focused on by this white paper. Some of these alternatives contain several of the advantages sought by fair value proponents, but at a smaller cost (in resources, subjectivity.) Each alternative also brings its own disadvantages, hence

there is no clear "right" answer. The selection of any financial accounting paradigm is at least partially a value judgement, not a pure scientific exercise.

- ***Reflection of credit standing is a controversial issue*** – There are arguments for and against the reflection of credit standing in fair value estimates of insurance liabilities. The task force has consciously avoided taking a position on this issue. Instead we have attempted to present both sides in a clear, objective fashion.

The task force chair wishes to thank all involved with this project for the tremendous amount of work done in a short period of time. In approximately six months, the task force team (with the help of key contributors) produced what I believe to be an excellent workproduct, one that hopefully will be a major contribution to the profession's understanding of the fair value issue. Thank you, once again.

Casualty Actuarial Task Force on Fair Value Liabilities
December 1999 - August 2000 members

Ralph Blanchard (chair)
Bob Butsic
Catherine Cresswell
Louise Francis
Aaron Halpert
Phil Heckman
Gerry Kirschner
Sarah Krutov
Mike McCarter
Gary Nickerson
Stewart Sawyer
Ernest Wilson
Bryan Young

(other contributors: David Appel, Paul Brehm, Roger Hayne, Gary Josephson, Joe Lebens, Steve Lowe, Glenn Meyers, Elizabeth Smith, Pat Teufel, Gary Venter, American Academy of Actuaries - Committee on Property and Liability Financial Reporting)

CAS Task Force on Fair Value Liabilities
White Paper on Fair Valuing Property/Casualty Insurance Liabilities
Section K - Appendices

Table of Contents

Appendix 1: CAPM Method

This appendix presents an example of computing a risk-adjusted discount rate using CAPM.

In its simplest form, the approach used in Massachusetts assumes that the equity beta for insurance companies is a weighted average of an asset beta and an underwriting beta. The underwriting beta can therefore be backed into from the equity beta and the asset beta.

here

β_e is the equity beta for insurance companies, or alternatively for an individual insurer
β_A is the beta for insurance company assets
β_u is the beta for insurance company underwriting profits
k is the funds generating coefficient, and represents the lag between the receipt of premium and the average payout of losses in a given line
s is a leverage ratio

Since

$$\beta_e = \frac{\mathrm{Cov}(r_e, r_M)}{\mathrm{Var}(r_M)}$$

or the equity beta is the covariance between the company's stock return and the overall market return divided by the variance of the overall market return. It can be measured by regressing historical P&C insurance company stock returns on a return index such as the S&P 500 Index. Similarly, β_A can be measured by evaluating the mix of investments in insurance company portfolios. The beta for each asset category, such as corporate bonds, stocks, real estate is determined. The overall asset beta is a weighted average of the betas of the individual assets, where the weights are the market values of the assets.

Example:
Assume detailed research using computerized tapes of security returns such as those available from CRISP concluded that β_e for the insurance industry is 1.0 and β_A for the insurance industry is 0.15. By examining company premium and loss cash flow patterns, it has been determined that k is 2. The leverage ratio s is assumed to equal 2. The underwriting Beta is

$$\beta_u = \frac{\beta_e - (ks+1)\beta_A}{s}$$

or $\beta_U = .5*(1. - (2*2+1).15) = .125$

Once β_U has been determined overall for the P&C industry, an approach to deriving the beta for a particular line is to assume that the only factor affecting the covariance of a given line's losses with the market is the duration of its liabilities:

$$\beta_L = -k\beta_U$$

So if the average duration in a given line is 2, its beta is -2*.125= -.25

In order to derive the risk-adjusted rate, the risk free rate and the market risk premium are needed. Assume the current risk free rate is 6% and the market risk premium (i.e., the excess of the market return over the risk free return) is 9%. Then the risk-adjusted rate is:

$$r_L = r_f + \beta_L(r_m - r_f)$$

or $r_L = .06 - .25 * (.09) = .06 - .0225 = .0375$

An alternative approach to computing the underwriting beta is to regress accounting underwriting returns in a line of business on stock market returns. The method suffers from the weakness that the reported underwriting returns often contain values for the liabilities that have been smoothed over the underwriting cycle, thus depressing their variability.

Appendix 2: IRR Method

All balance sheet values are at fair value. Thus, the liability value at each evaluation date must be calculated using a risk-adjusted interest rate. Since we are trying to find this value, it is an input that is iterated until the IRR equals the desired ROE. (This is easily done using the "Goal Seek" function in an Excel spreadsheet.)

The present value of the income taxes is a liability under a true economic valuation method. However, in the FASB and IASC proposals, it is not included.[1] The basis for this calculation is found in Butsic (Butsic, 2000). To a close approximation, the PV of income taxes equals the present value of the tax on investment income from capital, divided by 1 minus the tax rate. The PV is taken at an after-tax risk-free rate.

Exhibit A2 shows an example of the risk adjustment calculation, using the IRR method, for a liability whose payments extend for three periods.

[1] Note that the present value of income taxes is not the same as the deferred tax liability. For example, the present value of income taxes includes the PV of taxes on *future* underwriting and investment income generated by the policy cash flows.

Exhibit A2
Calculation of Risk Adjustment Using Internal Rate of Return Model

	Fixed Inputs		**Loss & LAE cash flow patterns**	
1	Risk-free rate	0.060		
2	Expected investment return	0.080		Proportion
3	Income tax rate	0.350	Time	of Total
4	Equity beta	0.800	0	0.000
5	market risk premium	0.090	1	0.500
6	Capital/reserve	0.500	2	0.300
7	Loss & LAE	1000.00	3	0.200
8			Total	1.000
9	**Calculated values**			
10	Required ROE	0.1320		
11	Risk-adjusted yield	0.0346		
12	After-tax risk-free rate	0.0390		
13	Premium	968.75		
14				
15	**Iterative Input**			
16	Risk adjustment	0.0254		
17				

		Time			
18	**Balance sheet, at fair value**				
19		**0**	**1**	**2**	**3**
20	*Assets*				
21	Investments, before dividend	960.14	1018.12	469.03	110.55
22	Investments, after dividend	1432.21	725.53	292.97	0.00
23					
24	*Liabilities*				
25	Loss & LAE	944.15	476.81	193.31	0.00
26	Income tax liability	24.60	10.31	3.01	0.00
27	Capital, before dividend	0.00	531.00	272.71	110.55
28	Capital after div (required amount)	472.07	238.41	96.66	0.00
29					
30	**Income**				
31	Underwriting income	24.60	-32.67	-16.50	-6.69
32	Investment income		114.58	58.04	23.44
33	Net income, pretax	24.60	81.91	41.54	16.75
34	Inv income, capital (risk-adjusted)		28.32	14.30	5.80
35					
36	**Insurance Cash Flows**				
37	Premium	968.75	0.00	0.00	0.00
38	Loss & LAE	0.00	-500.00	-300.00	-200.00
39	Income tax	-8.61	-28.67	-14.54	-5.86
40					
41	Income tax, capital (risk-adjusted)		9.91	5.01	2.03
42					
43	**Capital flow (dividend)**	472.07	-292.59	-176.06	-110.55
44					
45	**Internal rate of return**	13.20%			

Notes to Exhibit A2

Rows (Note that "R1" denotes Row 1, "R2" denotes Row 2, etc.):

1. Rate for portfolio of U. S. Treasury securities having same expected cash flows as the losses.
2. Expected return for the insurer's investment portfolio. Note that the yield on a bond is not an expected return. The yield must by adjusted to eliminate expected default. Municipal bond yields are adjusted to reflect the implied return as if they were fully taxable.
3. Statutory income tax rate on taxable income.
4. Estimates can be obtained from Value Line, Yahoo Finance or other services.
5. Estimates are commonly available in rate filings (e.g., Massachusetts).
6. All-lines value an be estimated by adjusting historical industry reserve values to present value and adding back the after-tax discount to GAAP equity. See Butsic (1999) for an example. For individual lines, a capital allocation method can be used, such as Myers and Read (1999).
7. An arbitrary round number used to illustrate the method.

10. R1 + (R4 x R5).
11. R1 – R16.
12. (1 – R3) x R1
13. R25 + R26 (at time 0).

16. This value is iterated until the IRR (Row 45) equals R10.

21. R22 (Prior Year) + R37 + R38 + R39.
22. R21 + R43.

25. Present value of negative R38 using interest rate R11.
26. Present value of R41 using interest rate R12. Result is divided by (1 – R3).
27. (R6, capital/reserve) x R25.
28. R27 + R43.

31. Time 0: R37 – R25. Time 1 to 3: – R11 x R25 (Prior Year).
32. (R22, Prior Year) x R2.
33. R31 + R32.
34. (R28, Prior Year) x R1.

37. R13.
38. – R7 x payment pattern in Rows 4 though 7.
39. – R3 x R33.

41. R3 x R34.

43. R28 – R27

45. Internal rate of return on Row 43 cash flows.

Appendix 3: Single Period RAD model

All balance sheet values are at fair value.

The discussion of the income tax liability is the same as in Appendix 2.

Here, there is no iteration needed, since the risk adjustment is derived directly from the equations relating the variables to each other. Butsic (2000) derives this result.

The formula is

$$z = c\left[\frac{R - r_f}{1 - t}\right] + (r_A - r_f)\left[1 + c\frac{1 + r_f}{1 + r_f(1 - t)}\right],$$

where the variables are:

z	risk adjustment to the risk-free rate
c	capital as a ratio to the fair value of the liability
R	required rate of return on capital (ROE)
r_A	expected return on assets (includes bond yields net of expected default)
r_f	risk-free rate
t	income tax rate

Although the risk adjustment can be calculated directly from the above formula, we have provided Exhibit A3, which shows that the risk adjustment in fact produces the required ROE and internal rate of return. The format of Exhibit A3 is similar to that of Exhibit A2. However, only a single time period is needed.

Note that exhibits A2 and A3 give slightly different results for the risk adjustment. This is because capital is needed for both asset and liability risk. In a multiple period model, the relationship between the assets and loss reserve fair value is not strictly proportional. This creates a small discrepancy.

Exhibit A3
Calculation of Risk Adjustment Using Single Period ROE Model

	Fixed Inputs		
1	Risk-free rate	0.060	
2	Expected investment return	0.080	
3	Income tax rate	0.350	
4	Equity beta	0.800	
5	market risk premium	0.090	
6	Capital/reserve	0.500	
7	Loss & LAE	1000.00	
8			
9	**Calculated values**		
10	Required ROE	0.1320	
11	Risk-adjusted yield	0.0348	
12	After-tax risk-free rate	0.0390	
13			
14	Premium	981.38	
15	Risk adjustment	0.02518	
16			
17	**Balance sheet, at fair value**	Time	
18		**0**	**1**
19	*Assets*		
20	Investments, before dividend	976.12	546.96
21	Investments, after dividend	1459.30	0.00
22			
23	*Liabilities*		
24	Loss & LAE	966.35	0.00
25	Income tax liability	15.02	0.00
26	Capital, before dividend	0.00	546.96
27	Capital after div (required amount)	483.18	0.00
28			
29	**Income**		
30	Underwriting income	15.02	-33.65
31	Investment income		116.74
32	Net income, pretax	15.02	83.10
33	Inv income, capital (risk-adjusted)		28.99
34			
35	**Insurance Cash Flows**		
36	Premium	981.38	0.00
37	Loss & LAE	0.00	-1000.00
38	Income tax	-5.26	-29.08
39			
40	Income tax, capital (risk-adjusted)		10.15
41			
42	**Capital flow (dividend)**	483.18	-546.96
43			
44	**ROE**		**13.20%**
45			
46	**Internal rate of return**	**13.20%**	

Notes to Exhibit A3

Rows (Note that "R1" denotes Row 1, "R2" denotes Row 2, etc.):

1. Rate for portfolio of U. S. Treasury securities having same expected cash flows as the losses.
2. Expected return for the insurer's investment portfolio. Note that the yield on a bond is not an expected return. The yield must by adjusted to eliminate expected default. Municipal bond yields are adjusted to reflect the implied return as if they were fully taxable.
3. Statutory income tax rate on taxable income.
4. Estimates can be obtained from Value Line, Yahoo Finance or other services.
5. Estimates are commonly available in rate filings (e.g., Massachusetts).
6. All-lines value an be estimated by adjusting historical industry reserve values to present value and adding back the after-tax discount to GAAP equity. See Butsic (1999) for an example. For individual lines, a capital allocation method can be used, such as Myers and Read (1999).
7. An arbitrary round number used to illustrate the method.

10. R1 + (R4 x R5).
11. R1 – R15
12. (1 – R3) x R1

14. R24 + R25 (at time 0).
15. R6 x (R10 – R1) / (1 – R3) – (R2 – R1) x [1 + R6 x (1 + R1) / (1+R12)].

20. R21 (Prior Year) + R36 + R37 + R38.
21. R20 + R42.

24. Present value of R7 using interest rate R11.
25. Present value of R40 using interest rate R12. Result is divided by (1 – R3).
26. Time 0: 0; Time 1: R20 – R24 – R25.
27. R6 x R24.

30. Time 0: R36 – R24. Time 1: – R11 x R24 (Prior Year).
31. (R21, Prior Year) x R2.
32. R30 + R31.
33. (R27, Prior Year) x R1.

36. R14.
37. Time 0: 0. Time 1: – R7.
38. – R3 x R32.

40. R3 x R33.

42. R27 – R26

44. (R26, Time 1) / (R27, Time 0) – 1.

46. Internal rate of return on Row 42 cash flows.

Appendix 4: Using Underwriting Data

This appendix describes Butsic's procedure for computing risk adjusted discount rates. The following relationship is used for the computation.

$$C = P(1+i)^{-u} - E(1+i)^{-w} - L(1+i_A)^{-t}$$

Where:

C is the cash flow on a policy and can be thought of as the present value of the profits, both underwriting and investment income, on the policy,
P is the policy premium,
E is expenses and dividends on the policy,
L is the losses and adjustment expenses,
$\underline{u}$ is the average duration of the premium, or the average lag between the inception of the policy and the collection of premium,
w is the average duration of the expenses,
t is the average duration of the liabilities.
i is the risk free rate of return
i_A is the risk adjusted rate of return

This formula says that the present value cash flow or present value profit on a group of policies is equal to the present value of the premium minus the present value of the components of expenses minus the present value of losses. Premiums and expenses are discounted at the risk free rate. Each item is discounted for a time period equal to its duration, or the time difference between inception of the policy or accident period and expiration of all cash flows associated with the item. Losses are discounted at the risk-adjusted rate. Underwriting data in ratio form, i.e., expense ratios, loss ratios, etc. can be plugged into the formula. When that is done, P enters the formula as 1, since the ratios are to premium.

In ratio form this formula would be:

$$c = 1(1+i)^{-u} - e(1+i)^{-w} - l(1+i_A)^{-t}$$

c is the ratio of present value profit to premium
e is the expense ratio, including dividends to policyholder
l is the loss ratio

Using as a starting point the rate of return on surplus, where the surplus supporting a group of policies is assumed to be eV_m, or the leverage ratio times the average discounted reserve, Butsic (Bustic, 1988) derived the following simplified expression for the risk adjustment:

$$Z = e(R-i) = (1+i)C/V_m ,$$

where:
Z is the risk adjustment to the interest rate or the percentage amount to be subtracted from the risk free rate = $e(R - i)$
C and i are as defined above
V_m is the average discounted reserve for the period

V_m is generally taken as the average of the discounted unpaid liabilities at the beginning of the accident or policy period (typically 100% of the policy losses) and the discounted unpaid liabilities at the end of the period. In general, this would be equal to 100% plus the percentage of losses unpaid at the end of the period (one year if annual data is used) divided by 2. The discount rate is the risk-adjusted rate. If V_m is computed as a ratio to premium, then published loss ratios are discounted and used in the denominator.

To complete the calculation, the quantity c, or the ratio of discounted profit to premium should be multiplied by $(1 + i)$ and divided by v_m (V_m in ratio form). To derive initial estimates of the risk adjustment, it is necessary to start with a guess as to the value of the risk adjustment to the discount rate in order to obtain a value for discounted liabilities.

The following is an example of the computation of the risk adjustment using this method. It is necessary to start with a guess for the risk adjustment and then perform the calculation iteratively until it converges on a solution. This example is based on data in Butsic's (1988) paper.

Parameter assumptions	
Interest Rate R_f	0.0972
Fraction of losses OS after 1 year	0.591
Initial Risk Adjustment	0.044

Variable	Nominal Value	Duration	Discounted Value
1 Loss&LAE	0.767	2.300	0.681
2 Premium	1.000	0.250	0.977
3 UW Expense	0.268	0.250	0.262
4 Pol Dividends	0.016	2.250	0.013
5 Average Liabilities	**0.610**	**1.800**	**0.556**

Calculation	
6 Premium-Expenses Discounted	
(2) - (3) - (4)	0.702
7 Premiums-Expenses-Losses Disc	0.021
(6)-(1)	
8 C*(1+I)	0.024
(7)*(1+I)	
9 Z=C*(1+I)/V_m	0.042
(8)/(5)	

An additive risk load

An additive or dollar risk load can be computed from the same data. The formula for the computation of a risk load is:

$$c = p(1+i)^{-u} - e(1+i)^{-w} - l(1+i)^{-t}$$

$$rl = c / l(1+i)^{-t}$$

Where rl is the additive risk load and i is the risk free interest rate.
An example is shown below:

Parameter assumptions	
Interest Rate Rf	0.0972

Variable	Nominal Value	Duration	Discounted Value
1 Loss&LAE	0.767	2.300	0.620
2 Premium	1.000	0.250	0.977
3 UW Expense	0.268	0.250	0.262
4 Pol Dividends	0.016	2.250	0.013

Calculation	
5 Premium-Expenses Discounted	
(2) - (3) - (4)	0.702
6 C =Premiums-Expenses-Losses Disc	0.083
(5)-(1)	
7 C/PV(Losses)	0.133
(6)/(1)	

Appendix 5: The Tax Effect

More recent work by Butsic (Butsic, 2000) has examined the effect of taxes on the risk adjusted discount rates and insurance premium. Butsic argued that, due to double taxation of corporate income, there is a tax effect from stockholder supplied funds. Stockholder funds are the equity supplied by the stockholder to support the policy. In the formulas above, stockholder supplied funds are denoted by E and taken to be the ratio of e to the present value of losses $V = L(1+i_A)^{-T}$. For a one period policy an amount E is invested at the risk free rate i, an amount E_i of income is earned, but because it is taxed at the rate t, the after tax income is $E_i(1 - t)$. The reduced investment income on equity will be insufficient to supply the amount needed to achieve the target return. In order for the company to earn its target after tax return, the amount lost to taxes must be included in the premium. However, the underwriting profit on this amount will also be taxed. The amount that must be added to premium to compensate for this tax effect is:

$$\frac{Eit}{(1-t)\left[1+i(1-t)\right]}$$

This is the tax effect for a one period policy if the discount rate for taxes is the same as the discount rate for pricing the policy, i.e., the risk adjusted discount rate. Butsic shows that there is an additional tax effect under the current tax law, where losses are discounted at a higher rate than the risk adjusted rate. There is also a premium collection tax effect, due to lags between the writing and collecting of premium. This is because some premium is taxed before it is collected. Butsic developed an approximation for all of these effects taken together, as well as the multiperiod nature of cash flows into the following adjustment to the risk adjusted discount rate:

$i_A' = i - e(1-t)(r_T - i)$,where
i_A' is the tax and risk adjusted rate,
e is a leverage ratio,
t is the tax rate,
r_T is the pre tax return on equity.

This is the effective rate used to discount losses to derive economic premium. The tax effect acts like an addition to the pure risk adjustment. Since premiums as stated in aggregate industry data already reflect this tax effect, no adjustment is needed for the risk adjusted discount rate used for pricing. However, for discounting liabilities, it may be desirable to segregate the tax adjustment from the pure risk adjustment, since the tax effect really represents a separate tax liability. Using the formula above, as well as the formula for determining the pure risk adjustment to the discount rate the two effects could be segregated. One would need to have an estimate of the total pre tax return on equity.

Appendix 6: Using Aggregate Probability Distributions

This example uses the Collective Risk Model to compute a risk load. It represents only one of the many approaches based on aggregate probability distributions. This is in order to keep the illustration simple.

The approach is based on the following model for risk load:

- Risk Load = λ SD[Loss] or Risk Load =λ Var[Loss],

Therefore, in order or compute a risk load, two quantities are needed: λ and Var[Loss], since SD(Loss) = Var[Loss]$^{1/2}$. The following algorithm from Meyers (Meyers, 1994) will be used to compute the variance of aggregate losses.

The Model:

1. Assume claim volume has an unconditional Poisson distribution.

2. Assume the Poisson parameter, n (the claim distribution mean), varies from risk to risk.

3. Select a random variable χ from a distribution with mean 1 and variance c.

4. Select the claim count, K, at random from a Poisson distribution with mean χn, where the random variable χ is multiplied by the random Poisson mean n.

<u>The Variability of Insurer Losses</u>

5. Select occurrence severities, $Z_1, Z_2, .., Z_K$, at random from a distribution with mean μ and variance σ^2.

6. The total loss is given by:

$$X = \sum_{j=1}^{K} Z_j$$

The expected occurrence count is n (i.e. E[(χ n] = E[n] = n). n is used as a measure of exposure.
When there is no parameter uncertainty in the claim count distribution $c = 0$,

$$\text{Var}[x] = n\,(\mu^2 + \sigma^2),$$

and variance is a linear function of exposures.

When there is parameter uncertainty:

$$\text{Var}[x] = nu + n^2v,$$

where

$$u = (\mu^2 + \sigma^2)$$

and

$$v = c\mu^2$$

nu is the process risk and n^2v is the parameter risk.

For example, assume an insurer writes two lines of business. The expected claim volume for the first line is 10,000 and the expected claim volume for the second line is 20,000. The parameter c for the first line is 0.01 and for the second line is 0.005. Let the severity for line 1 be lognormal with a mean of $10,000 and volatility parameter (the standard deviation of the logs of losses) equal to 1.25 and the severity for line 2 be lognormal with severity of $20,000 and volatility equal to 2. Applying the formula above for the variance of aggregate losses, we find that the variance for line 1 is 1.05×10^{14} and the variance of line two is 1.24×10^{15} and the sum of the variances for the two lines is 1.34×10^{15}. The standard deviation is $36,627,257.

One approach to determining the multiplier λ would be to select the multiplier ISO uses in its increased limits rate filings. In the increased limits rate filings, λ is applied to the variance of losses and is on the order of 10^{-7}.(Meyers, 1998)

In recent actuarial literature, the probability of ruin has been used to determine the multipliers of SD(loss) or Var(Loss). (Kreps 1998, Meyers 1998, Philbrick, 1994). The probability of ruin or expected policyholder deficit is used to compute the amount of surplus required to support the liabilities. To keep the illustration simple, we use the probability of ruin approach. However, the expected policyholder deficit or tail value at risk (which is similar to expected policyholder deficit) approaches better reflect the current literature on computing risk loads. Suppose the company wishes to be 99.9% sure that it has sufficient surplus to pay the liabilities, ignoring investment income, the company will require surplus of 3.1 times the standard deviation of losses, if one assumes that losses are normally distributed.[2] In order to complete the calculation, we need to know the company's required return on equity, r_e. This can be determined by examining historical return data for the P&C insurance industry. Then the required risk margin for one year is $r_e \times 3.1 \times 36{,}627{,}257$. For instance, if r_e is 10% then the risk margin is

[2] If one assumes that aggregate losses are lognormally distributed, then the company needs approximately $e^{(2.33*.06)}*$ the expected losses as surplus, where .06 is the volatility parameter, derived from the mean and variance of the distribution..

11,354,450 or about 2.0% of expected losses. In this example, the parameter lambda is equal to 3.1 r_e. The result computed above could be converted into a risk margin for discounted losses by applying the 2% to losses discounted at the risk free rate. This would require the assumption that the risks of investment income on the assets supporting the losses being less than expected is much less than the risk that losses will be greater than expected. When the assets supporting the liabilities are primarily invested in high quality bonds, this assumption is probably reasonable. (see D'Arcy et. al., 1997)

Philbrick in his paper commissioned by the CAS "Accounting for Risk Margins" had a slightly different approach to determining the risk margin. Philbrick's formula for risk margin, given a total surplus requirement S, (i.e. 3.1* standard deviation in this example), a rate of return on equity r_e and a risk free rate i is:

$$RM = \frac{(r_e - i) x S}{1 + r_e}$$

This is a risk margin for discounted losses not undiscounted losses. . The formula above assumes that some of the required return on surplus is obtained from investing the surplus at the risk free rate. If $i = 5\%$, and $r_e = 10\%$ the risk margin in this illustration would be \$5,161,113.

In this example, it should be noted that the majority of the standard deviation is due to parameter risk, as process risk for such large claim volumes is minimal. However, only parameter risk for claims volumes has been incorporated. A more complete model would incorporate parameter risk for the severity distribution. This risk parameter has been denoted the "mixing parameter" in the actuarial literature. The algorithm for incorporating this variance into the measure of aggregate loss variance is as follows:

1 - 5. Follow steps 1 through 5 above, describing the selection of frequency and severity parameters for a distribution

6. Select a random variable B from a distribution with mean 1 and variance b.

7. The total loss is given by:

$$X = \sum_{j=1}^{K} Z_j / \mathrm{B}$$

The variance reflecting the mixing parameter is given by:

$$\mathrm{Var}[x] = n(1+b)(\mu^2 + \sigma^2) + n^2(b + c + bc)\mu^2.$$

Procedures for estimating b and c are provided by Meyers and Schenker. The procedures use the means and variances of the claim count and the loss distribution to compute b and c. The parameter b can also be viewed as the uncertainty contributed to the total estimate of losses due to uncertainty in the trend and development factors. Methods for measuring the variance due to development are presented by Hayne, Venter and Mack. Regression statistics containing information about the variances of trend factors are published in ISO circulars and can be developed from internal data. To continue our example, we will assume that the b parameter for line 1 is 0.02 and for line 2 is 0.05. Then the standard deviation of aggregate losses is $95,663,174. The risk load using Philbrick's formula is $13,479,811 or 2.7% of expected undiscounted losses. The load is intended to be applied to <u>discounted</u> liabilities where liabilities are discounted at the risk free rate. Thus if losses take one year to pay out the risk margin is 2.8% of the present value of liabilities.

The above risk load is consistent with liabilities that expire in one year. When losses take more than one year to pay, Philbrick uses the following formula to derive a risk load.

$$RM = \sum_j \frac{(r_e - i)S_j}{(1+r_e)^j}$$

This formula can be applied to liabilities of any maturity. Where S_j is the surplus requirement for outstanding liabilities as of year j. In the above example if losses pay out evenly over 3 years then the risk margin is $20,693,737or 4.6% of he discounted liabilities. The calculation is shown below.

(1)	(2)	(3)	(4)	(5)
	Surplus			
t	.227*PV(OS Losses)	1/(1+r(e)^t	(3)*(2)	(r(e)-.05)*(4)
0	219,965,641	1.000	219,965,641	10,998,282
1	146,643,760	0.909	133,312,510	6,665,625
2	73,321,880	0.826	60,596,595	3,029,830
				20,693,737

The computation above assumes that the relative variability of the liabilities remains constant as the liabilities mature. As this may not be the case, refinements to the measure of variability by age of liability may be desirable. One approach to modeling the uncertainty in reserves would derive measures of variability from observed loss development variability. This is the approach used by Zenwirth, Mack and Hayne. Another approach, consistent with how risk base capital is computed, would measure historic reserve development for P&C companies for a line of business from Schedule P.

Appendix 7: Direct Estimation of Market Values

Below we illustrate how to estimate the risk adjustment to the interest rate for a single firm, based on empirical data.

Assume that the market value of assets is 1400 and the book (undiscounted) value of the liabilities is 1000. Both of these values are available from the insurer's published financial statements. Also, assume that using the Ronn-Verma method (see the discussion in the Credit Risk Appendix), the estimated market value of the firm's equity is 500 and that the value of the expected default (the credit risk adjustment) is 10. The market value of the equity adjusted to exclude default is 510.

The discounted risk adjusted liabilities equals the market value of the assets minus the market value of the equity or 900 = 1400 – 500. The implied market value of the liabilities adjusted for default equals the market value of the assets minus the market value of the equity adjusted for default, or **890** = 1400 – 510.

Assume that the risk-free interest rate applicable to valuing the insurer's expected liability payments is 6% and that the liability payment pattern is 10% per year for 10 years (paid at the end of each year). The present value of the liabilities at the risk-free rate is 730. Thus, the risk margin, expressed in dollars is **160** = 890 – 730. Alternatively, the interest rate that gives a present value of 890 using the above payment pattern is 2.18%. This value implies a risk adjustment of **3.82%**.

The following discussion provides an example of the Ronn and Verma method.

Let A be the market value of assets, L the market value of liabilities and σ the volatility of the asset/liability ratio. The formula for the owners' equity, where there is a possibility of default, is the call option with expiration in one year:

(1) $$E = A \cdot N(d) - L \cdot N(d - \sigma),$$

where $d = \ln(A/L)/\sigma + \sigma/2$ and $N(d)$ is the standard normal distribution evaluated at d.

Notice that equity value with no default is simply $E_n = A - L$. For an insurer with stochastic assets and liabilities, σ_E, the volatility of the equity, is related to the asset/liability volatility by

(2) $$\sigma_E = N(d)A\sigma / E.$$

Equations (1) and (2) are solved simultaneously to get E and σ.

The expected default value equals $E - E_n$, or the derived market value of the equity minus the equity value with no default.

The method is easily demonstrated with a numerical example. Assume that A =130, L = 100 and σ_E = 0.5. Solving the simultaneous equations gives E = 40.057 and σ = 0.117. Therefore, the value of the expected default is

0.057 = 40.057 – 40.000.

For an insurer, the market value of assets is readily determined from the published balance sheet. Discounting the reserves at a risk-free rate can approximate the market value of liabilities. The equity volatility can be estimated by analyzing the insurer's stock price over a recent time frame, as done by Allen, Cummins and Phillips.

Appendix 8: Distribution Transform Method

Assume expected claim counts for a policy equal 100 and ground up severities follow a Pareto distribution:

$$F(x) = 1 - [b/(b+x)]^q \text{ for x>0.}$$

Therefore $G(x) = [b/(b+x)]^q$

$E[X] = b/(q-1)$

E(aggregate loss) = 100* E[X]

For the transformation r, $G(x) = [b/(b+x)]^{qr}$.

If the market risk premium is 10% then risk loaded premiums equal:

$$100\frac{b}{q-1}1.1 = 100\frac{b}{qr-1}$$

This expression can be solves for r:

$$r = [(q-1)/1.1+1]/q = (q+0.1)/1.1q.$$

If q were 2, r would be 0.95.

Expected values for higher layers could be computed by replacing q with qr in the Pareto distribution and using the Pareto formula for limited expected value to price the excess layers.:

E(X, x) = Limited Expected Value function = $E[X]\{1-[b/(b+x)]^{q-1}\}$

In the above example, a transformation was applied only to the severity distribution. However, with a little more work, the transformation could be applied to both the frequency and severity distribution.

For instance the formula for the transformed mean of a Poisson distribution with a mean of 100 and transformation parameter r is:

$$\sum_{j}((e^{200}\Gamma(j)-\Gamma(j,100))/\Gamma(j))^{r}$$

This formula could be combined with the formula for the transformed severity distribution to produce a risk loaded mean.

Appendix 9: Credit Risk

The Time Horizon Problem

In general, long-tail liabilities are subject to greater default risk than are short-tail liabilities. To see why this is so, assume that an ongoing insurer has a 1% chance of insolvency each year. The insurer has two lines of business: line A has claims that are paid in one year and line B has claims that are paid in five years. The probability of a claim from line A not being paid in full is 1%. Assuming that each year's insolvency potential is independent of the other years, the probability of a claim from line B not being fully paid is $4.9\% = 1-(1-0.01)^5$, or about 5 times as great as for line A.

An insurance firm's owners normally make capital decisions at an approximate annual frequency, so to truly measure the long-term value of the potential default, it is necessary to consider the future capital flows as well as the current level of capital. (However, note that the fair value accounting proposals purposely ignore future transactions that are not based on current contractual obligations.) The complexity due to future capital flows (which are contingent on future company results and market conditions) makes the estimation of credit risk extremely difficult.

To make the credit risk adjustment calculation more tractable, it is customary to assume an annual time horizon and that future insolvencies have the same probability as for the current one-year horizon. For longer-term liabilities, one can further assume that the insolvency probabilities are independent year-to year and then determine the overall expected default by a formula suggested by the above 5-year calculation:

$$D = \frac{D_1}{p}[1 - w_1(1-p) - w_2(1-p)^2 - \ldots - w_n(1-p)^n] \quad \cong \quad D_1[w_1 + 2w_2 + \ldots + nw_n].$$

Here, D_1 is the fair value of the expected default for the one-year horizon, p is the one-year insolvency probability and the weight w_i is the expected proportion of loss paid in year i (the weights sum to 1). Using the approximation above, the fair value over an n year time horizon of a company's option to default can be expressed as a function of its one year default value.

It should be noted that the published research relating to bond default rates does not support the assumption that annual default rates over the life of a bond are independent and identically distributed. That is, for many categories of bonds, the default rate during the third and fourth year is higher than the default rate during the first and second years after issuance. If the assumption of independent and identically distributed default rates is inappropriate for bonds, it may be inappropriate for some of the companies issuing bonds (i.e. insurance companies) and therefore the approximations in the above formula would not be appropriate.

A related technical issue that must be addressed in calculating the credit risk adjustment is the length of the *time horizon* over which defaults are recognized. At one extreme, it may be argued the applicable horizon is unlimited. Insurers are obliged to pay claims occurring during the contractual coverage period, no matter how long the reporting and settlement processes take. On the other hand, solvency monitoring and financial reports have a quarterly or annual cycle. Also, it is important to recognize that capital funding and withdrawal decisions are made with an approximate quarterly or annual cycle. An approach that often makes the solution easier to derive is to assume that one may view the time horizon as being a fairly short duration. According to this view, if the company is examined over short increments such as one year, corrective action is applied and insolvency over a longer term is avoided. The task force considers this view to be controversial. The alternative view is that insurance liabilities are often obligations with relatively long time horizons, and these longer horizons need to be considered when evaluating the companies' option to default on its obligations.

In the numerical examples below, we have determined the annual fair value of default. The extension to longer-duration liabilities is straightforward, using the above formula, if one assumes the formula to be appropriate. If one assumes the formula to be inappropriate, many of the methods below can be modified to adjuste for the longer time horizon of insurance liabilities.

Numerical Examples of Credit Risk Adjustment Estimation Methods

1. Implied Option Value: Example

The following (until #2, the DFA example), is a repeat of a few pages ago immediately following Appendix 7,

The following discussion provides an example of the Ronn and Verma method.

Let A be the market value of assets, L the market value of liabilities and σ the volatility of the asset/liability ratio. The formula for the owners' equity, where there is a possibility of default, is the call option with expiration in one year:

$$(1) \qquad E = A \cdot N(d) - L \cdot N(d-\sigma),$$

where $d = \ln(A/L)/\sigma + \sigma/2$ and $N(d)$ is the standard normal distribution evaluated at d.

Notice that equity value with no default is simply $E_n = A - L$. For an insurer with stochastic assets and liabilities, σ_E, the volatility of the equity, is related to the asset/liability volatility by

(2) $\sigma_E = N(d) A\sigma / E$.

Equations (1) and (2) are solved simultaneously to get E and σ.

The expected default value equals $E - E_n$, or the derived market value of the equity minus the equity value with no default.

The method is easily demonstrated with a numerical example. Assume that A =130, L = 100 and σ_E = 0.5. Solving the simultaneous equations gives E = 40.057 and σ = 0.117. Therefore, the value of the expected default is

0.057 = 40.057 - 40.000.

For an insurer, the market value of assets is readily determined from the published balance sheet. Discounting the reserves at a risk-free rate can approximate the market value of liabilities. The equity volatility can be estimated by analyzing the insurer's stock price over a recent time frame, as done by Allen, Cummins and Phillips.

2. Dynamic Financial Analysis: Example

An insurer has initial liabilities of $100 million, measured at fair value, but under the assumption that all contractual obligations will be paid. Assume that the DFA model has been run using 10,000 simulations. The time horizon is one year. We examine all observations where the terminal fair value (before default) of liabilities exceeds the market value of the assets. Suppose that there are 22 of them, with a total deficit (liability minus asset value) of $660 million. The average default amount per simulation is $0.066 million.

The expected terminal fair value is then discounted at a risk-adjusted interest rate to get the fair value of the credit risk adjustment. With a 4% risk-adjusted interest rate, for example, the fair value of the default is **$0.063 million** = 0.066/1.04. Thus, the fair value of the liabilities, adjusted for credit risk, is **$99.94 million** ($100 million - $.06 million.

3. Rating Agency Method: Example

This example shows how the table of default ratios might look, if a one-year time horizon approach was used. Alternatively, a matrix of default ratios by rating and lag year could be used, similar to those available from Moody's (e.g., Moody's January 2000 report titled "Historical Default Rates of Corporate Bond Issuers, 1920-1999"). Here the ratings are the current A. M. Best categories. The values in the table below are purely hypothetical.

Rating	Annual Expected Default Ratio (Raw Results)	Annual Expected Default Ratio (Adjusted)
A++	0.000%	0.001%
A+	0.000%	0.004%
A	0.013%	0.010%
A-	0.043%	0.050%
B++	0.122%	0.100%
B+	0.155%	0.150%
B	0.432%	0.300%
B-	0.619%	0.500%
C++	0.653%	0.800%
C+	1.221%	1.000%
C	1.554%	1.500%
C-	2.221%	2.000%
D	4.689%	5.000%
E	13.658%	15.000%

The raw results would be based on historical insolvency data. A simulation model or a closed-form model could be applied to a large sample of companies within each rating group to produce the adjusted results. These results might be further adjusted to ensure that a higher rating had a corresponding lower default expectation.

To show how the above table would be applied, assume that an insurer has initial liabilities of $100 million. These are measured at fair value, but under the assumption that all contractual obligations will be paid. Assume also that the insurer has an A- Best's rating. The expected default is 0.05% of $100 million, or $50,000.

The expected terminal fair value is then discounted at a risk-adjusted interest rate to get the fair value of the credit risk adjustment. With a 4% risk-adjusted interest rate, for example, the fair value of the default is **$48,100** = 50,000/1.04. Thus, the fair value of the liabilities, adjusted for credit risk, is **$99.95 million**.

Appendix 10: References

Section A - Background

1) FASB Preliminary Views document titled "Reporting Financial Instruments and Certain Related Assets and Liabilities at Fair Value", available for download (via the "Exposure Drafts" link) at:
http://www.rutgers.edu/Accounting/raw/fasb/new/index.html

2) FASB Project Update - Fair Value - FASB web site at:
http://www.rutgers.edu/Accounting/raw/fasb/new/index.html

3) FASB Statement of Financial Accounting Concepts No. 7 - "Using Cash Flow Information and Present Value in Accounting Measurements", available for purchase from the FASB at:
http://www.rutgers.edu/Accounting/raw/fasb/public/index.html

4) IASC Issues Paper on Insurance, available for download at:
http://www.iasc.org.uk/frame/cen3_113.htm

5) Philbrick, Stephen W., "Accounting for Risk Margins," Casualty Actuarial Society Forum, Spring 1994, Volume 1, pp. 1-87, available for download at:
http://www.casact.org/library/reserves/94SPF1.PDF

Section D - Methods of Estimating Fair Value

Conceptual overview - risk margins

1) Stulz, Rene, "Whats wrong with modern capital budgeting?", Address to the Eastern Finance Association, April, 1999

Method 1 - The CAPM Approach

1) Automobile Insurance Bureau of Massachusetts, 1998 Massachusetts Private Passenger Automobile Underwriting Profit Filing

2) D'Arcy, S. P., and Doherty, N. A., "The Financial Theory of Pricing Property-Liability Insurance Contracts," Huebner Foundation, 1988

3) Fairley, William, "Investment Income and Profit Margins in Property Liability Insurance: theory and Empirical Evidence," Fair Rate of Return in Property-Liability Insurance, Cummins, J.D., Harrington S.A., Eds, Kluwer-Nijhoff Publishing, 1987, pp. 1-26.

4) Fama, Eugene and French, Kenneth, "The Cross Section of Expected Stock Returns" Journal of Finance, Vol 47, 1992, pp. 427-465

5) Fama, Eugene and Kenneth French, "Industry Costs of Equity," Journal of Financial Economics, Vol 43, 1997, pp. 153 - 193

6) Feldblum, Shalom, "Risk Load for Insurers", PCAS LXXVII, 1990, pp. 160- 195

7) Kozik, Thomas, "Underwriting Betas-The Shadows of Ghosts," PCAS LXXXI, 1994, pp. 303-329.

8) Mahler, Howard, "The Meyers-Cohn Profit Model, A Practical Application," PCAS LXXXV, 1998, pp. 689 – 774.

9) Meyers, S and Cohn, R, "A Discounted Cash Flow Approach to Property-Liability Rate Regulation," Fair Rate of Return in Property-Liability Insurance, Cummins, J.D., Harrington S.A., Eds, Kluwer-Nijhoff Publishing, 1987, pp. 55-78.

10) Myers, S. C. and R. Cohn, 1987, "Insurance rate of Return Regulation and the Capital Asset Pricing Model, Fair Rate of Return in Property Liability Insurance", in J. D. Cummins and S. Harrington, eds. Kluwer-Nijhoff Publishing Company, Norwell MA.

Method 2 & 3 - The Pricing-Based Methods

1) Brealy, Richard A. and Stuart C. Myers, 1996, "Principles of Corporate Finance (5th Edition)", McGraw-Hill, New York.

2) Cummins, J. David, "Multi-Period Discounted Cash Flow Ratemaking Models in Property-Liability Insurance," Journal of Risk and Insurance, March 1990, Volume 57:1, pp. 79-109.

3) Roth, R., "Analysis of Surplus and Rates of Return Using Leverage Ratios", 1992 Casualty Actuarial Society Discussion Paper Program - Insurer Financial Solvency, Volume 1, pp 439-464

Method 3 - The Single-period RAD (Risk-Adjusted Discount) method

1) Butsic, Robert, "Determining the Proper Discount Rate for Loss Reserve Discounting: An Economic Approach," 1988 Casualty Actuarial Society Discussion Paper Program - Evaluating Insurance Company Liabilities, pp. 147-188.

2) D'Arcy, Stephen P., 1988, "Use of the CAPM to Discount Property-Liability Loss Reserves", Journal of Risk and Insurance, September 1988, Volume 55:3, pp. 481-490.

Method 4 - Methods Based on Underwriting Data

1) Butsic, Robert, "Determining the Proper Discount Rate for Loss Reserve Discounting: An Economic Approach," 1988 Casualty Actuarial Society Discussion Paper Program - Evaluating Insurance Company Liabilities, pp. 147-188.

2) Butsic, Robert P., 2000, Treatment of Income Taxes in Present Value Models of Property-Liability Insurance, Unpublished Working Paper.

3) Myers, S and Cohn, R, "A Discounted Cash Flow Approach to Property-Liability Rate Regulation," Fair Rate of Return in Property-Liability Insurance, Cummins, J.D., Harrington S.A., Eds, Kluwer-Nijhoff Publishing, 1987, pp. 55-78

Method 5 - Actuarial Distribution-Based Risk Loads

1) Butsic, Robert P., 1994, "Solvency Measurement for Property-Liability Risk-Based Capital Applications", Journal of Risk and Insurance, 61: 656-690.

2) Cornell, Bradford, "Risk. Duration and Capital Budgeting: New Evidence on Some Old Questions", Journal of Business, 1999 vol 72, pp 183-200.

3) Hayne, Roger M., "Application of Collective Risk Theory to Estimate Variability in Loss Reserves," Proceedings of the Casualty Actuarial Society (PCAS), LXXVI, 1989, p. 77-110

4) Heckman, Philip, "Seriatim, Claim Valuation from Detailed Process Models," paper presented at Casualty Loss Reserve Seminar, 1999.

5) Heckman, Philip and Meyers, Glenn, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," PCAS, 1983, pp. 22-621

6) Kreps, Rodney, "Investment Equivalent Reinsurance Pricing," PCAS, 1998

7) Kreps, Rodney E., "Reinsurer Risk Loads from Marginal Surplus Requirements," Proceedings of the Casualty Actuarial Society (PCAS), LXXVII, 1990, p. 196

8) Mack, Thomas, "Which Stochastic Model is Underlying the Chain Ladder Method,", CAS Forum, Fall 1995, pp.229-240

9) Meyers, Glenn, "The Cost of Financing Insurance", paper presented to the NAIC's Insurance Securitization Working Group at the March 2000 NAIC quarterly meeting.

10) Meyers, Glenn G., "The Competitive Market Equilibrium Risk Load Formula for Increased Limits Ratemaking," Proceedings of the Casualty Actuarial Society (PCAS), LXXVIII, 1991, pp 163-200

11) Meyers, Glenn G., "Risk Theoretic Issues in Loss Reserving: The Case of Workers Compensation Pension Reserves," Proceedings of the Casualty Actuarial Society (PCAS), LXXVI, 1989, p. 171

12) Meyers, Glen and Schenker, Nathaniel, "Parameter Uncertainty in the Collective Risk Model," PCAS, 1983, pp. 111-143

13) Philbrick, Stephen W., "Accounting for Risk Margins," Casualty Actuarial Society Forum, Spring 1994, Volume 1, pp. 1-87

14) Stulz, Rene, "Whats wrong with modern capital budgeting?", Address to the Eastern Finance Association, April, 1999

15) Zehnwirth, Ben, "Probabilistic Development Factor Models with Application to Loss Reserve Variability, Prediction Intervals and Risk Based Capital," CAS Forum, Spring 1994, pp. 447-606.

Method 7 - Direct estimation of market values

1) Allen, Franklin, J. David Cummins and Richard D. Phillips, 1998, "Financial Pricing of Insurance in a Multiple Line Insurance Company", Journal of Risk and Insurance, 1998, volume 65, pp. 597-636.

2) Ronn, Ehun I., and Avinash K. Verma, 1986, Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model, Journal of Finance, 41(4): 871-895.

Method 8 - Distribution Transform Method

1) Butsic, Robert P, 1999, Capital Allocation for Property Liability Insurers: A Catastrophe Reinsurance Application. Casualty Actuarial Society Forum, Fall 1999.

2) Venter, Gary G., 1991, Premium Implications of Reinsurance Without Arbitrage, ASTIN Bulletin, 21 No. 2: 223-232.

3) Venter, Gary G., 1998, (Discusssion of) Implementation of the PH-Transform in Ratemaking, [by Shaun Wang; presented at the Fall, 1998 meeting of the Casualty Actuarial Society]

4) Wang, Shaun, 1998, Implementation of the PH-Transform in Ratemaking, [Presented at the Fall, 1998 meeting of the Casualty Actuarial Society].

Miscellaneous "methods" references

1) Derrig, Richard A., 1994, Theoretical Considerations of the Effect of Federal Income Taxes on Investment Income in Property-Liability Ratemaking, Journal of Risk and Insurance, 61: 691-709.

2) Meyers, Glenn G., "The Cost of Financing Catastrophe Insurance," Casualty Actuarial Society Forum - Summer 1998, pp. 119 - 148.

3) Meyers, Glenn G., "Calculation of Risk Margin Levels for Loss Reserves," 1994 Casualty Loss Reserve Seminar Transcript

4) Myers, S. C. and J. Read, 1998, "Line-by-Line Surplus Requirements for Insurance Companies," [Unpublished paper originally prepared for the Automobile Insurance Bureau of Massachusetts.]

5) Robbin, Ira, The Underwriting Profit Provision, CAS Study Note, 1992

Section H - Credit Standing and Fair Value Liabilities

Method 1 - Implied Option Value

1) Allen, Franklin, J. David Cummins and Richard D. Phillips, 1998, "Financial Pricing of Insurance in a Multiple Line Insurance Company", Journal of Risk and Insurance, 1998, volume 65, pp. 597-636.

2) Black, Fischer and Myron Scholes, 1973, The pricing of Options and Corporate Liabilities, Journal of Political Economy, May-June, 81: 637-659.

3) Butsic, Robert P., 1994, "Solvency Measurement for Property-Liability Risk-Based Capital Applications", Journal of Risk and Insurance, 61: 656-690.

4) Cummins, J. David, 1988, Risk-Based Premiums for Insurance Guaranty Funds, Journal of Finance, September, 43: 823-838

5) Derrig, Richard A., 1989, Solvency Levels and Risk Loadings Appropriate for Fully Guaranteed Property-Liability Insurance Contracts: A Financial View, Financial Models of Insurance Solvency, J. D. Cummins and R. A. Derrig eds., Kluwer Academic Publishers, Boston, 303-354.

6) Doherty, Neil A. and James R. Garven, 1986, Price Regulation in Property-Liability Insurance: A Contingent-Claims Approach, Journal of Finance, December, 41: 1031-1050.

7) Ronn, Ehun I., and Avinash K. Verma, 1986, Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model, Journal of Finance, 41(4): 871-895.

Method 2 - Stochastic modeling using Dynamic Financial Analysis (DFA)

1) CAS Valuation and Financial Analysis Committee, Subcommittee on Dynamic Financial Models, "Dynamic Financial Models of Property/Casualty Insurers," CAS Forum, Fall 1995, pp. 93-127.

2) Correnti, S.; Sonlin, S.M.; and Isaac, D.B., "Applying A DFA Model to Improve Strategic Business Decisions," CAS Forum, Summer 1998, pp. 15-51.

3) D'Arcy, Stephen P.; Gorvett, Richard W.; Herbers, Joseph A.; Hettinger, Thomas E.; Lehmann, Steven G.; and Miller, Michael, "Building a Public Access PC Based DFA Model," Casualty Actuarial Society Forum, Summer 1997, Vol. 2, pp.1-40

4) D'Arcy, S.P.; Gorvett, R.W.; Hettinger, T.E.; and Walling, R. J., "Using the Public Access DFA Model: A Case Study," CAS Forum, Summer 1998, pp. 53-118.

5) Kirschner, G.S.; and Scheel, W.C., "Specifying the Functional Parameters of a Corporate Financial Model for Dynamic Financial Analysis," CAS Forum, Summer 1997, Volume 2, pp. 41-88. Although the candidate should be familiar with the information and concepts presented in the exhibits, no questions will be drawn directly from them.

6) Lowe, S.P.; and Stanard, J.N., "An Integrated Dynamic Financial Analysis and Decision Support System for a Property Catastrophe Reinsurer," ASTIN Bulletin, Volume 27, Number 2, November 1997, pp. 339-371.

Method 3 - Incorporate historic default histories by credit rating from public rating agencies.

1) Altman, Edward, "Measuring Corporate bond Mortality and Performance", The Journal of Finance, Sept, 1989, pp.909-922

Determining the Change in Mean Duration Due to a Shift in the Hazard Rate Function

Daniel R. Corro

Abstract:

From a major world event (such as a military action) to a seemingly minor detail (such as the use of a new plastic washer in a faucet design) change must be accounted for when collecting, interpreting and analyzing data. Indeed, the intervention itself may be the focus of the study. Theoretically, the best way to model some interventions, especially time-dependent ones, is via the hazard function. On the other hand, it may be necessary to translate into simpler concepts in order to answer practical questions. The average duration, for example, may have well-understood relationships with costs, making it the best choice for presenting the result.

For example, Shuan Wang [3] discusses deforming the hazard function by a constant multiplicative factor—proportional hazard transform—as a way to price risk load, with the mean playing the role of the pure loss premium.

This paper investigates how a shift in the hazard rate impacts the mean. The primary focus of the discussion is the case of bounded hazard rate functions of finite support. A formal framework is defined for that case and a practical calculation is described for measuring the impact on the mean duration of any deformation of the hazard function. The primary tool is the Cox Proportional Hazard model. Several formal results are derived and concrete illustrations of the calculation are provided in an Appendix, using the SAS implementation. The paper establishes that the method can be applied in a very general context and, in particular, to deformations which are not globally proportional shifts. Indeed, the method demands no assumed form for either the survival distribution or the deformation. The discussion begins with a case study that illustrates the application of these ideas to assess the cost impact of a TPA referral program.

Introduction

Recall that the survival function, $S(t)$, is just the probability of surviving to maturity time t and that the hazard function , $h(t)$, is the rate of failure at time t. We assume some general familiarity with these concepts in this discussion—they are introduced formally in Section II. While both functions equally well determine a model of survivorship, the survival function is the more common and the hazard function the more arcane. Often though, the best way to model a change in circumstances, especially a time-dependent intervention, is via the hazard function. On the other hand, it may be necessary to translate into simpler concepts in order to answer practical questions. The average duration, for example, may have a well-understood relationship with costs which makes it the best choice for presenting the result.

For example, Shuan Wang [3] discusses deforming the hazard function by a constant multiplicative factor—proportional hazard transform—as a way to price risk load, with the mean playing the role of the pure loss premium.

This paper investigates how a shift in the hazard rate impacts the mean. The primary focus of the discussion is the case of bounded hazard rate functions of finite support. A formal framework is defined for that case and a practical calculation is described for measuring the impact on the mean duration of any deformation of the hazard function. The primary tool is the Cox Proportional Hazard model (see [1]). Several formal results are derived and concrete illustrations of the calculation are provided in an Appendix, using the Statistical Analysis System [SAS] implementation (c.f. [1]) of the Cox model. The paper establishes that the method can be applied in a very general context and, in particular, to deformations which are not globally proportional shifts. Indeed, the method demands no assumed form for either the survival distribution or the deformation.

The paper begins with a case study that illustrates how these ideas were used to assess the cost impact of a Third Party Administrator (TPA) referral program. While this paper has a distinctly theoretical focus, the best way to explain the basic concepts is through a real world example. Indeed, most of the ideas are a direct consequence of attempts to achieve a better understanding of the case study outlined in Section I. The study illustrates that for most practical issues it is sufficient to determine the mean duration to failure via numerical integration. For many purposes, there is little need to invoke the more esoteric results developed in the subsequent sections. Still, the example illustrates the potential value of building a survivorship model whose hazard structure is designed to accommodate the issues under consideration. Among the technical results of the paper is a description of just such a survivorship model. While the discussion of the case study is largely self-contained for anyone generally familiar with the terminology of survivorship models, the discussion does make an occasional reference to the notation and observations developed in the subsequent sections.

Section II introduces the notation and formal set-up. The language shifts from rather discursive to decidedly technical. Section III discusses some well known examples. The remainder of the paper is devoted to several technical findings on how duration is impacted by a hazard shift. Specifically, Section IV discusses the case of finite support that is the case of primary interest. Section V considers how to combine hazards of finite support into more complex models suited to empirical data and the kind of investigation described in the case study.

Section I: A Case Study

Consider the following situation (while the data is based on a real world study, some liberties are taken in this discussion; in particular, the thought process, as described, follows hindsight more than foresight). The context is workers compensation (WC) insurance. We are required to assess whether a third party claims administrator (TPA) is saving money for two of its clients that have been selectively referring a portion of their WC claims over to be managed by the TPA. These clients are both large multi-state employers that are "self-insured" inasmuch as they do not purchase a WC insurance policy. The medical bills and loss of wages benefits are the direct responsibility of the employers and each has built internal systems to process their WC claims. The data captured by these systems is designed for administering claims, however, rather than for analytical use. As such, the data is comparatively crude relative to claim data of insurance companies or TPA's. They do, however, capture the date and jurisdiction of the injury, a summary of payments made to date, as well as if and when the claim is settled. There are, however, no "case reserves" available nor are there sufficient details, such as impairment rating or diagnosis, to adequately assess the severity of the claim.

Over the past few years, the employers have selectively farmed out the more complex claims to the TPA. The TPA has its own claim data on the cases referred to it and there is sufficient overlap to identify common claims within the TPA and the employer files. Moreover, the TPA files are more like insurance carrier data files and contain considerably more information, including the date of the referral, impairment ratings, claimant demographics and other claim characteristics.

A major problem is referral selection bias. The selection process itself is not well defined, even within an employer. Also, when the TPA first entered the picture, a greater percentage of referred claims were older, outstanding cases. Simply comparing the average cost per case of referred versus retained cases would not yield any meaningful information. Indeed, the selection process refers claims that are more expensive. Not only does this result in a higher severity for the referred cases, it renders the retained cases less severe over time. In such a circumstance, no matter how successful the TPA is in reducing costs, its mean cost per case will be comparatively high.

One fact that stood out for both employers is that the percentage of cases that closed within one year had more than doubled since the TPA became involved. Also, the referral rate shot up dramatically, suggesting that the TPA is, at some pragmatic level, viewed as being effective. Of course, that could also be the effect of imposed cost reductions on the staff the employer is now willing to maintain for WC claims handling, given the money spent on the TPA.

Another complicating issue is that the benefits that will be paid on some WC claims are paid out over many years. Without any consistent reserves it is very problematic to find comparable data. The challenge here is to make an assessment using the currently available payment data.

Without the presence of case reserves or enough claim characteristics to grade the severity of the claims, conventional actuarial approaches do not work well. As noted, the employer data, being collected largely for administrative purposes, did include the key dates of injury and settlement. This, combined with what was noted in regard to claim closure rates, suggested an approach based on survival analysis. In this context, a "life" corresponds with a claim, beginning at the date of injury and "failing" at claim settlement. Information on unsettled (open) claims is then "right censored". It was hoped that the survival analysis models would enable us to deal with censored data, since there were no case reserves available for that purpose.

Merging the TPA data together with the employer claim data, we built a data set that included an indicator of referral and, where so indicated, the date of referral. Other covariates captured are:

Explanatory Variables Used in the Proportional Hazard Duration Models	
Description	**Variable Name(s)**
Indicator of which of the two employers the claimant worked for	EMPL2
Indicators of the year of the injury (year 1992 as base)	AY93,AY94
Indicator whether a medical fee schedule applies in the state of jurisdiction	MF01
Indicator whether employer choice of physician applies in the state of jurisdiction	EC01
Indicator whether the nature of the injury is a sprain or strain (subjective)	NOI_SPR
Indicator whether the nature of the injury is a cut or laceration (objective)	NOI_CUT
Indicator whether the claim was referred to the TPA	TPA
Time dependent indicator whether the claim was referred to the TPA	TxTPA = 0 prior to TPA referral TxTPA =1 after TPA referral. The x refers to 3 time frames of referral from date of injury: x=1 within 1^{st} 6 months x=2 within 2^{nd} 6 months x=3 after 1 year

A claim survivorship model was constructed from this data. As defined in later sections in a formal way, the conceptual base of the model is a "hazard" function. The model assumes that the various explanatory variables impact the hazard function as a proportional shift, i.e., multiplication by a constant proportionality factor. Such survivorship models are referred to as proportional hazard models. Referral to the TPA is an exception in as much as it is captured as a so-called "time-dependent" intervention.

Instead of a constant value for the explanatory variable, the TPA referral indicator is allowed to take on two values so as to be able to capture into the model the time frame of referral (=0 prior to referral, =1 afterward). The proportional adjustment factor associated with TPA referral confirmed the expectation that referral was associated with a greater hazard, i.e., shortened claim duration. While the effect on the hazard was measured, the assignment demanded that it be translated into savings. In order to do that, it was necessary to convert the result back into factors related to claim costs. Whence the basic question of this paper: how to translate a change in hazard into a change in (mean) duration.

The task is to assess the cost impact of the TPA program, but that is not clearly defined. Due to the limited time frame of the data, the lack of case reserves or multiple loss valuations, it was clear that the "ultimate" cost impact could not be assessed using the available data, at least not directly. Also, "ultimate cost impact" is a more complicated notion than what the clients were after. We interpreted the task more simply: since we had the actual payments made on TPA referred cases, what we needed to measure is hypothetical: what would the payments on those claims have been without use of the TPA?

There is a catch, however. Consider a simplified case: the "original" payout pattern is $1 per day for 100 days on all claims. Assume that the referral to the TPA results in a single $100 payment on the first day. A little thought will convince the reader that at any point in time, ignoring discounting and the prospect that the business fails, the TPA will appear more costly. The comparison will not be fair unless it takes into account the unpaid balance: no matter how simply you frame the issue, reserves cannot be completely ignored.

The data included payment and duration, so there were ways available to translate a change in mean duration to dollars. Our choice was to use the non-referred claims to build a regression model in which the dependent variable is (log of) the benefits paid to date. The explanatory variables would include available claim characteristics together with the (log of) the payment duration. The characteristics (such as employer, accident year, jurisdiction or nature of injury, as above, together with perhaps additional covariates if available like age, wage, gender, part of body) are assumed independent of TPA referral and their mean values over the TPA-referred claims are readily determined. The only missing piece is the duration variable. Again, the question reduces to the topic of this paper: determining the impact on the mean duration.

The Cox proportional hazard model is well suited to this context. The model was run on pooled TPA-referred and non-referred data, with TPA-referral included among the explanatory variables in the model. This captures TPA-referral as a deformation of the hazard function and the methods of the paper can be applied to finish the job. Appendix 2 provides output that details the calculations.

The case study, however, illustrates an additional complexity. More precisely, the TPA-referral was incorporated into the Cox proportional hazard model as a time-dependent intervention (both the date of injury and date of referral being available). Also, as in the paper, the deformation of the hazard function was modeled as a combination of proportional shifts over three time intervals, as shown in the following table (refer to Appendix 2, page 5 of the listing):

Time Period I	**Hazard Ratio φ_i**
1st 6 months	1.424
2nd 6 months	1.203
After 1 year	1.122

The pattern of the hazard ratios supports the TPA's contention that its early intervention is more cost effective. Indeed, TPA intervention has its greatest, and most statistically significant, impact during the first six months. Although not critical to this context, that was an important finding of the study.

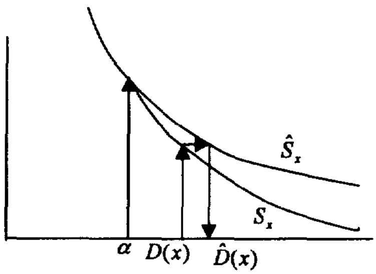
The difference in the values of the hazard ratios suggests that not only is it appropriate to model TPA referral as a time-dependent intervention, it is also appropriate to mitigate the global proportional hazard assumption by specializing to several time intervals. This is a very direct approach to that issue; the technical discussion of the subsequent sections follows that approach. An alternative way to mitigate that assumption—the one in fact used in the study report—is to group the TPA intervention by the lag time to referral. That formulation produces similar results and more directly supports the greater impact of early intervention. Conceptually, it is easy to regard TPA-referral within a few days of the injury as being an essentially different intervention than referral after several months.

The remainder of this discussion is somewhat more technical and makes reference to some of the notation and results presented in the subsequent sections of the paper.

The SAS PHREG procedure is used not only to estimate the three proportional hazard ratios φ_i. It optionally outputs paired values $(t, S(t))$ of a "baseline" survival function $S(t)$ at time t as well. We chose to determine a baseline survival function, $S(t)$, corresponding to the value of 0 for all covariates in the model. In particular, it applies to the case of non-referral as defined by the vanishing of the TPA-referral indicator variable. Observe that for the purpose of determining the baseline survival, only the non-time dependent TPA-referral indicator is used, since the baseline option is not available in the presence of time-dependent interventions.

This baseline survival function provides the expected duration distribution for the non-referred claims at the formal value 0 for the other explanatory variables in the model. Because referral is captured as a time dependent intervention, the deformation of the hazard function is itself dependent on the lag time to referral of the individual claims. Consequently, no single survival function of the form $S_\delta(t)$ (see Section II) can suffice to measure the impact on mean duration. This presents a somewhat more complicated situation than that considered in this paper.

To deal with this, let x represent a TPA-referred claim and $\beta = \beta(x)$ be the proportional hazard ratio associated to x by the model, which therefore includes the factor $\varphi = \varphi(x)$ for the TPA-referral as a time-dependent intervention. Let $D(x)$ represent the claim duration function; recall that we seek a hypothetical alternative $\hat{D}(x)$ which associates what the duration would have been had x not been referred. Letting $S_x, \hat{S}_x$ denote, respectively, the survival curves for x with and without referral, and $\alpha = \alpha(x)$ the lag time to TPA-referral, we have the following picture:



The idea is adjust duration so as to hold "maturity" constant. It follows from observations in Section II that:

$$\hat{S}_x(t) \approx S(t)^{\frac{\beta}{\varphi}} \quad \text{and} \quad S_x(t) = \begin{cases} \hat{S}_x(t) & t \le \alpha \\ \hat{S}_x(t)^{\varphi}\hat{S}_x(\alpha)^{1-\varphi} & t \ge \alpha \end{cases}$$

It follows, taking the $\left(\frac{\beta}{\varphi}\right)^{th}$ root, that:

$$S_x(D(x)) \approx \hat{S}_x(\hat{D}(x)) \quad \Leftrightarrow \quad S(\hat{D}(x)) \approx \left(S(D(x))\right)^{\varphi} S(\alpha)^{1-\varphi}$$

Since the baseline survival curve $S(t)$ is known, this provides a way to determine $\hat{D}(x)$ for any TPA-referred claim x. The methods described in the paper can now be invoked to estimate what the mean payment duration of those claims would have been had they not been referred. Again, the details of the calculation can be found in Appendix 2. The following table summarizes the findings in the case study (pages 10 and 15 of the listing):

Assumption	Mean Duration
As Referred to TPA (actual)	0.737 years
No Referral (hypothetical)	0.826 years

Note that the application of the logic used to define $D(x) \mapsto \hat{D}(x)$ becomes somewhat problematic when crossing a boundary of the time intervals used to define the φ_i. That is another reason that, in the study, we chose to partition the TPA-intervention by layer of referral lag $\alpha = \alpha(x)$.

Finally, these mean duration figures can be plugged into the cost models and translated into dollar savings attributable to TPA-referral. This case study is included to illustrate a non-traditional application of survival analysis to an insurance problem, emphasizing the power that manipulating the hazard function can bring to the analysis. The remainder of the paper develops a formal context in which this can be done. The focus is on formal relationships between the more "arcane" changes in hazard and the more "presentable" effect on mean duration.

Section II: Basic Terminology and Notation

Let $\Re^+$ denote the set of nonnegative real numbers. Let $h(t)$ denote a function from some subinterval $\Sigma \subseteq \Re^+$ to $\Re^+$. The set Σ is called the *support.* We assume throughout that $h(t)$ is (Lebesgue) integrable on Σ and that $0 \in \overline{\Sigma}$ is in the closure of the support. Any such $h:\underset{t \mapsto h(t)}{\Sigma \to \Re^+}$ can be viewed as a hazard rate function and survival analysis associates the following three functions:

$$g:\underset{t \mapsto g(t)}{\Sigma \to \Re^+} \text{ where } g(t) = \int_0^t h(s)ds$$

$$S:\underset{t \mapsto S(t)}{\Sigma \to [0,1]} \text{ where } S(t) = e^{-g(t)}$$

$$f:\underset{t \mapsto f(t)}{\Sigma \to \Re^+} \text{ where } f(t) = -\frac{dS}{dt} = h(t)S(t)$$

As is customary, we refer to $S(t)$ as the *survival function,* $f(t)$ as the *probability density function [PDF]* and t as time. We also let T denote the random variable for the distribution of survival times and $\mu = E_T(T)$ the mean duration. When we adorn $h(t)$ with a subscript, superscript, etc., we make the convention that these associated functions all follow suit. There are many well-known relationships and interpretations of these functions—refer to Allison[1] for a particularly succinct discussion which also discusses the SAS implementation of the Cox proportional hazard model.

Provided f is differentiable at t, it is readily determined whether the hazard rate is increasing or decreasing at t:

$$\frac{dh}{dt} = \frac{\frac{df}{dt}}{S} + h^2;\ h \text{ is decreasing at } t \Leftrightarrow \frac{df}{dt} < -\frac{f^2}{S}$$

In particular, it is a necessary—but by no means sufficient—condition that the density be decreasing in order for the hazard to be decreasing.

We are concerned with what happens when $h(t)$ is changed or "shifted" in some fashion. This paper deals particularly with proportional shifts as the Cox model provides a viable way to measure that type of shift (c.f.[2]). More precisely, we are interested in shifts of the form:

$$\delta = \delta(\alpha,\varphi) \text{ for } \alpha,\varphi \geq 0 \text{ where } \delta(h) = h_\delta \text{ is defined as } h_\delta(t) = \begin{cases} h(t) & t \leq \alpha \\ \varphi h(t) & t > \alpha \end{cases}$$

The following are immediate consequences of this definition and our notational conventions:

$$g_\delta(t) = \begin{cases} g(t) & t \le \alpha \\ g(\alpha) + \varphi(g(t) - g(\alpha)) \\ = (1-\varphi)g(\alpha) + \varphi g(t) & t \ge \alpha \end{cases}$$

$$S_\delta(t) = \begin{cases} S(t) & t \le \alpha \\ S(\alpha)^{1-\varphi} S(t)^{\varphi} & t \ge \alpha \end{cases}$$

$$f_\delta(t) = \begin{cases} f(t) & t \le \alpha \\ \varphi\left[\dfrac{S(t)}{S(\alpha)}\right]^{\varphi-1} f(t) & t \ge \alpha \end{cases}$$

We are particularly interested in the effect that such a shift has on mean duration, which is formally captured in the function:

$$\Delta(h;\alpha,\varphi) : \underset{\mu}{\Re^+} \underset{\mapsto}{\rightarrow} \underset{\mu_{\delta(\alpha,\varphi)}}{\Re^+}$$

While at first these shifts may seem restrictive, one of the main results of this paper is to show that the ability to measure these shifts is sufficient for handling very general problems. In fact, it will be shown that even when dealing with time-dependent interventions one can generally make do with the ability to handle the case $\alpha = 0$, in which case we make the common identification $\delta(0,\varphi) = \varphi$ of scalar multiplication with the scalar itself. Accordingly, we have

$$h_\varphi(t) = \varphi h(t),\ S_\varphi(t) = S(t)^\varphi \quad \textit{and} \quad f_\varphi(t) = \varphi S(t)^{\varphi-1} f(t) \quad \textit{for all } \varphi \ge 0, t \in \Sigma$$

Section III illustrates this notation in the case of two of the (infinite support) distributions commonly used in survival analysis. However, we choose to deal exclusively with the case of hazard functions with finite support in the remainder of the paper. Section IV discusses the additional assumptions, notation and conventions applicable specifically to finite support hazard functions and presents some examples. Section V discusses decomposing and combining finite support hazards and presents the main result: a formula for calculating the effect on mean duration of a shift in the hazard rate function. We also provide two appendices that detail the calculations referenced in the paper using SAS and, in particular, illustrate how the SAS proportional hazards model procedure (PHREG) can be used to do all the heavy lifting.

Section III: Familiar Examples

In this section we illustrate our notation with some distributions with infinite support $\Sigma = (0,\infty)$ which have found common application in survival analysis. The first three are selected to present straightforward illustrations of the notation and concepts and for those we only consider the case $\alpha = 0$ (recall the identification $\delta(0,\varphi) = \varphi$). We begin with the simplest example of a hazard function:

Example III.1. *Constant hazard function*: Let $h(t) \equiv 1$, then:

$$h_\varphi(t) = \varphi \quad g_\varphi(t) = \varphi t \quad S_\varphi(t) = e^{-\varphi t} \quad f_\varphi(t) = \varphi e^{-\varphi t}$$

and a straightforward integration by parts yields $\Delta(0,\varphi) = \mu_\varphi = \int_0^\infty \varphi t e^{-\varphi t} dt = \frac{1}{\varphi}$.

Example III.2. *Increasing hazard function*: Let $h(t) = t$, then:

$$h_\varphi(t) = \varphi t \quad g_\varphi(t) = \frac{\varphi t^2}{2} \quad S_\varphi(t) = e^{-\frac{\varphi t^2}{2}} \quad f_\varphi(t) = \varphi t e^{-\frac{\varphi t^2}{2}}$$

The motivated reader may readily verify, via another integration by parts and exploiting the symmetry of the normal PDF, that:

$$\Delta(0,\varphi) = \mu_\varphi = \int_0^\infty \varphi t^2 e^{-\frac{\varphi t^2}{2}} dt = \sqrt{\frac{\pi}{2\varphi}}$$

Example III.3. *Decreasing hazard function*: Let $h(t) = \frac{1}{1+t}$, then:

$$h_\varphi(t) = \frac{\varphi}{1+t} \quad g_\varphi(t) = \varphi \ln(1+t) \quad S_\varphi(t) = \frac{1}{(1+t)^\varphi} \quad f_\varphi(t) = \frac{\varphi}{(1+t)^{\varphi+1}}$$

In this case, integration by parts together with l'Hospital's rule gives:

$$\Delta(0,\varphi) = \mu_\varphi = \int_0^\infty \frac{\varphi t}{(1+t)^{\varphi+1}} dt = \int_0^\infty \frac{dt}{(1+t)^\varphi} - \lim_{t\to\infty} \frac{1}{\varphi(1+t)^{\varphi-1}}$$

in which the right hand side limits both exist for $\varphi > 1$. For $\varphi = 1$ the right hand side diverges to $+\infty$, whence $\mu_\varphi \geq \mu_1$ is infinite for $\varphi \leq 1$. This illustrates that a proportional increase in the hazard function can reduce an infinite mean duration to a finite number and, conversely, that a proportional decrease can make a finite mean duration become infinite.

The next example describes one of the most popular survival distributions, often defined via its PDF:

Example III.4. *Weibulll density with parameters* $a,b > 0$. In this example, define

$$f(a,b;t) = abt^{b-1}e^{-at^b}$$

then (see, e.g. [2] Hogg-Klugman, pp. 231-232)

$$S(t) = e^{-at^b}; \; h(t) = abt^{b-1}; \quad \text{and} \quad \mu = \frac{\Gamma\left(\frac{1}{b}\right)}{ba^{\frac{1}{b}}}$$

This distribution conforms to a proportional hazard model, indeed:

$$f_\varphi(a,b;t) = f(\varphi a,b;t),$$

$$S_\varphi(t) = e^{-\varphi a t^b}, \; h_\varphi(t) = \varphi abt^{b-1} \quad and \quad \mu_\varphi = \frac{\Gamma\left(\frac{1}{b}\right)}{b(\varphi a)^{\frac{1}{b}}} = \frac{\mu}{\varphi^{\frac{1}{b}}}$$

Letting $\Gamma(\alpha)\Gamma(\alpha;t) = \int_0^t s^{\alpha-1}e^{-s}ds$ define the incomplete gamma function (as in [2], p. 217), we leave to the reader the verification that for the Weibull density:

$$\Delta(a,b;\alpha,\varphi) = \mu_{\delta(\alpha,\varphi)}$$

$$= a^{-\frac{1}{b}}\Gamma\left(\frac{b+1}{b}\right)\left[\Gamma\left(\frac{b+1}{b};a\alpha^b\right) + e^{a\alpha^b(\varphi-1)}\varphi^{-\frac{1}{b}}\left[1-\Gamma\left(\frac{b+1}{b};\varphi a\alpha^b\right)\right]\right]$$

When $\alpha = 0, a = b = 1$ this reduces to Example III.1; when $\alpha = 0, a = \frac{1}{2}, b = 2$ this reduces to Example III.2.

Example III.5. *Pareto density with parameters* $a,b > 0$. In this example, define

$$f(a,b;t) = ab^a(b+t)^{-a-1}$$

then (see, e.g. [2], pp. 222-223)

$$S(t) = \left(\frac{b}{b+t}\right)^a, \; h(t) = \frac{a}{b+t} \quad \text{and for } a > 1 \quad \mu = \frac{b}{a-1}$$

This distribution conforms to a proportional hazard model, indeed:

$$f_\varphi(a,b;t) = f(\varphi a,b;t),$$

$$S_\varphi(t) = \left(\frac{b}{b+t}\right)^{\varphi a}, \; h_\varphi(t) = \frac{\varphi a}{b+t} \quad \text{and for } \varphi a > 1, \; \mu_\varphi = \frac{b}{\varphi a - 1}$$

We again leave to the reader the verification that for the Pareto denisty:

$$
\begin{aligned}
\Delta(a,b;\alpha,\varphi) &= \mu_{\delta(\alpha,\varphi)} \\
&= \frac{b}{a-1} - \left(\frac{b}{b+\alpha}\right)^{a}\left(\frac{b+\alpha}{a-1}\right) + \left(\frac{b}{b+\alpha}\right)^{\varphi a}\left(\frac{b+\alpha}{\varphi a-1}\right) \\
&\qquad + \alpha\left[\left(\frac{b}{b+\alpha}\right)^{\varphi a} - \left(\frac{b}{b+\alpha}\right)^{a}\right]
\end{aligned}
$$

When $\alpha = 0$, $a = b = 1$ this reduces to Example III.3.

The last two examples are suggestive of the common approach to performing calculations in survival analysis: first, we select a form for the distribution, then we fit parameters to the data. Finally, we calculate whatever statistics are needed using formulas specific to that distribution (e.g. as found in [2]). This paper suggests the expediency of a simpler more empirical approach to calculating μ_δ that avoids making any assumptions as to the form of the distribution as well as any parameter estimation. Also, we can use the method with time-dependent interventions and it is especially easy to do in practice.

Section IV: Hazard with Finite Support

Most survival analysis discussions use distributions whose natural support is the set of positive real numbers, as in the previous section. The impetus for this work came from insurance, particularly claims analysis. Although actuaries customarily employ the usual collection of survival distributions--with their infinite supports--in practical applications claim duration is subject to limits. Moreover, the specific structure of the very far "tail" is either intrinsically unknowable, irrelevant, or both. Accordingly, this study focuses on the situation in which the data is limited to a finite time interval.

As described in the case study section, the insurance problem prompting this investigation arose in the line of workers compensation insurance. A very small percentage of those claims involve pension benefits that can continue for decades. Even the best insurance data bases, however, rarely track a coherent set of losses for more than 10 annual evaluations. That study concerned the implementation of a new program and the available data consisted of a one snapshot evaluation of claims captured into various automated systems. The data typically went back only four years and even the most matured cohort included a high percentage of open ("right censored") cases.

In this section we introduce the assumptions and notation for our case of interest: support $\Sigma = (0,1]$. We make the assumption that $h(t)$ is piecewise continuous. Observe that $g(t)$ and $S(t)$ are both continuous on $[0,1]$, the former nondecreasing and the latter nonincreasing. Let $p = S(1), 0 \le p \le 1$. The distribution T has a point mass of probability p at $\{1\}$. We will make extensive use of the following:

Proposition IV.1: *For any positive integer n:*

$$E(T^n) = n\int_0^1 t^{n-1}S(t)dt.$$

Proof: The proof is really just the integration by parts the diligent reader would have done a few times already in the previous section:

$$u = -S(t) \quad du = f(t)dt; \quad v = t^n \quad dv = nt^{n-1}dt$$

$$E(T^n) = \int_0^1 t^n f(t)dt + p = \int_0^1 v du + p = uv\Big]_0^1 - \int_0^1 u dv + p$$

$$= -t^n S(t)\Big]_0^1 + n\int_0^1 t^{n-1}S(t)dt + p = -p + n\int_0^1 t^{n-1}S(t)dt + p = n\int_0^1 t^{n-1}S(t)dt$$

completing the proof.

Letting σ^2 denote the variance of T, the following two corollaries are apparent:

Corollary IV.1:

i) $\mu = \int_0^1 S(t)dt$

ii) $\mu^2 + \sigma^2 = 2\int_0^1 tS(t)dt$.

Corollary IV.2:

$$\Delta(\alpha,\varphi) = \mu_\delta = \int_0^\alpha S(t)dt + S(\alpha)^{1-\varphi}\int_\alpha^1 S(t)^\varphi dt$$.

In particular, observe that $\mu_\varphi = \int_0^1 S(t)^\varphi dt$. It is intuitively clear that increasing the hazard decreases the mean duration, i.e., that $\Delta = \mu_\delta$ is a decreasing function of φ. A bit more thought should convince the reader that $\Delta = \mu_\delta$ is an increasing function of α for $\varphi > 1$ and decreasing for $\varphi < 1$. Since $g(t)$ is increasing, the following result formalizes this:

Proposition IV.2:

i) $\frac{\partial\Delta}{\partial\alpha} = (\varphi-1)f(\alpha)S(\alpha)^{-\varphi}\int_\alpha^1 S(t)^\varphi dt$

ii) $\frac{\partial\Delta}{\partial\varphi} = S(\alpha)^{1-\varphi}\left[g(\alpha)\int_\alpha^1 S(t)^\varphi dt - \int_\alpha^1 g(t)S(t)^\varphi dt\right]$.

Proof: i) From Corollary IV.2, the fundamental theorem of calculus and the product rule for differentiation:

$$\frac{\partial\Delta}{\partial\alpha} = \frac{\partial}{\alpha}\left[\int_0^\alpha S(t)dt - S(\alpha)^{1-\varphi}\int_1^\alpha S(t)^\varphi dt\right]$$

$$= S(\alpha) - \left[S(\alpha)^{1-\varphi}S(\alpha)^\varphi + \int_1^\alpha S(t)^\varphi dt\big((1-\varphi)S(\alpha)^{-\varphi}(-f(\alpha))\big)\right]$$

$$= S(\alpha) - S(\alpha) + (1-\varphi)S(\alpha)^{-\varphi}f(\alpha)\int_1^\alpha S(t)^\varphi dt$$

$$= (\varphi-1)S(\alpha)^{-\varphi}f(\alpha)\int_\alpha^1 S(t)^\varphi dt$$

ii) Observe that:

$$S_\delta(t) = \begin{cases} S(t) & t \le \alpha \\ S(\alpha)^{1-\varphi} S(t)^\varphi & t \ge \alpha \end{cases}$$

$$\Rightarrow \quad \frac{\partial S}{\partial \varphi} = \begin{cases} 0 & t \le \alpha \\ S(\alpha)^{1-\varphi}\left(\ln(S(t))S(t)^\varphi\right) & \\ -S(t)^\varphi\left(\ln(S(\alpha))S(\alpha)^{1-\varphi}\right) & t \ge \alpha \end{cases}$$

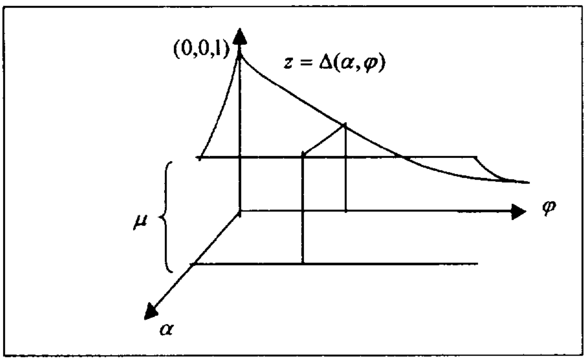
Noting that our assumptions enable us to differentiate under the integral, and recalling that $g(t) = -\ln(S(t))$, we find that:

$$\frac{\partial \Delta}{\partial \varphi} = \frac{\partial}{\partial \varphi}\left[\int_0^1 S_\delta(t)dt\right] = \int_0^1 \frac{\partial S_\delta(t)}{\partial \varphi}dt = -S(\alpha)^{1-\varphi}\left[\int_\alpha^1 S(t)^\varphi g(t)dt - g(\alpha)\int_\alpha^1 S(t)^\varphi dt\right]$$

which completes the proof.

The graph of $\Delta = \mu_\delta$ is a tent with a single "pole" of unit height at the origin, a front wall of infinite length and constant height μ and a back wall of decreasing height:

$$\Delta(0,0) = 1 \quad and \quad \forall \alpha, \varphi \ge 0, \quad \Delta(1,\varphi) = \Delta(\alpha,1) = \mu$$



Since $g(t)$ is nondecreasing, we clearly have:

$$-\ln(p)\mu = g(1)\mu \ge \int_0^1 g(t)S(t)dt \quad and \quad \mu = \int_0^1 tf(t)dt + p \ge p$$

The following refines this:

Lemma IV.1: $-\ln(p)\mu + p - \mu \geq \int_0^1 g(t)S(t)dt$

Proof: Set

$$u = g(t) \quad du = h(t)dt; \quad v = \int_0^t S(w)dw \quad dv = S(t)dt$$

$$\int_0^1 g(t)S(t)dt = g(t)\int_0^t S(w)dw\Big]_0^1 - \int_0^1 h(t)\int_0^t S(w)dwdt$$

$$= g(1)\mu - \int_0^1 f(t)\int_0^t \frac{S(w)}{S(t)}dwdt$$

$$\leq -\ln(p)\mu - \int_0^1 f(t)\int_0^t 1dwdt \quad as \quad \frac{S(w)}{S(t)} \geq 1 \ for \ w \leq t$$

$$= -\ln(p)\mu - \int_0^1 f(t)tdt = -\ln(p)\mu - (\mu - p).$$

Applying the lemma to the hazard function $h_\varphi(t)$:

$$0 \geq \frac{\partial \Delta}{\partial \varphi}\bigg|_{\alpha=0} = -\int_0^1 g(t)S(t)^\varphi dt \geq \frac{\mu_\varphi(1 + \varphi \ln(p)) - p^\varphi}{\varphi}.$$

which formally confirms how $\Delta(\alpha, \varphi)$ flattens as $\varphi \to \infty$. On the other hand, observe that if $h(a) > 0$ for some $a \geq 0$, then,

$$g(t) > 0 \ for \ t \geq a \Rightarrow S(t) < 1 \ \ for \ t \geq a$$

$$\Rightarrow \lim_{\varphi\to\infty} \mu_\varphi = \lim_{\varphi\to\infty} \Delta(0,\varphi) = \lim_{\varphi\to\infty} \int_0^1 S(t)^\varphi dt \leq \int_0^a dt = a$$

While the effect of an increase (decrease) of the hazard function clearly has the opposite affect on the mean duration, the effect on the variance is unclear. Indeed, the reader can use Corollary IV.1 to verify that:

$$\lim_{\varphi\to\infty} \sigma_\varphi = \lim_{\varphi\to 0} \sigma_\varphi = 0$$

Before we discuss some examples, we note the following integration formula, in which

$\mathrm{B}(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\quad \forall\, a,b>0$, the usual beta and gamma functions.

Lemma IV.2: For $a,c>-1, b>0$

$$\int_0^1 t^a(1-t^b)^c dt=\frac{\mathrm{B}\left(\frac{a+1}{b},c+1\right)}{b}$$

Proof: Letting $x=t^b \Rightarrow dx=bt^{b-1}dt$, then

$$\int_0^1 t^a(1-t^b)^c dt=\frac{1}{b}\int_0^1 t^{a-b+1}(1-t^b)^c bt^{(b-1)}dt$$

$$=\frac{1}{b}\int_0^1\left(x^{\frac{1}{b}}\right)^{a-b+1}(1-x)^c dx$$

$$=\frac{1}{b}\int_0^1 x^{\frac{a+1}{b}-1}(1-x)^c dx=\frac{\mathrm{B}\left(\frac{a+1}{b},c+1\right)}{b},$$

as claimed.

We next present some examples. The first, while especially simple, will play a major role in later findings.

Example IV.1. *Constant hazard function*, let $h(t)\equiv 1, 0\le t\le 1$. Then, as in Example III.1, we have:

$$h_\varphi(t)=\varphi;\quad g_\varphi(t)=\varphi t;\quad S_\varphi(t)=e^{-\varphi t};\quad f_\varphi(t)=\varphi e^{-\varphi t}$$

and we observe that

$$p_\varphi=S_\varphi(1)=e^{-\varphi}=p^\varphi,\quad \text{where } p=p_1=\frac{1}{e}$$

More generally, for $0\le\alpha\le 1$, we find:

$$h_\delta(t)=\begin{cases}1 & t\in[0,\alpha]\\ \varphi & t\in[\alpha,1]\end{cases}$$

$$g_\delta(t)=\begin{cases}t & t\in[0,\alpha]\\ \varphi(t-\alpha)+\alpha & t\in[\alpha,1]\end{cases}$$

$$S_\delta(t)=\begin{cases}e^{-t} & t\in[0,\alpha]\\ p^{\alpha}e^{-\varphi(t-\alpha)} & t\in[\alpha,1]\end{cases}$$

$$\Delta(\alpha,\varphi)=\mu_\delta=\int_0^1 S_\delta(t)dt=\int_0^\alpha e^{-t}dt+p^{\alpha}\int_\alpha^1 e^{-\varphi(t-\alpha)}dt=1-p^{\alpha}+\frac{p^{\alpha}}{\varphi}\left(1-p^{\varphi(1-\alpha)}\right)$$

In particular, we find that $\Delta(0,\varphi)=\mu_\varphi=\dfrac{1-p^{\varphi}}{\varphi}$. We will make considerable use of this example in later sections where we deal with combining hazards and show how to use the Cox Proportional Hazard model to approximate any hazard of finite support by a step function.

Example IV.2. *Increasing hazard function*, select $p\in[0,1]$, and define $f(t)=1-p, t\in[0,1]$; then:

$$S(t)=1-(1-p)t \quad and \quad h(t)=\frac{1-p}{1-(1-p)t}$$

This is an example of an increasing hazard that is not a proportional hazard model. We note that $h(t)$ is defined and continuous on $[0,1]$ for $p>0$, while the case $p=0$ is reminiscent of the infinite support case via the transformation $t\leftrightarrow\dfrac{t}{t+1}$. Finally, we note that the case $p=1\Rightarrow S(t)\equiv 1$ is of little interest, so we require $p<1$.

We leave to the reader the straightforward verification that in this case;

$$\Delta(\alpha,\varphi)=\mu_\delta=\alpha-\frac{(1-p)\alpha^2}{2}+\frac{p^{\varphi+1}(1-(1-p)\alpha)^{1-\varphi}-(1-(1-p)\alpha)^2}{(\varphi+1)(p-1)}$$

In particular,

$$\mu_\varphi=\frac{1-p^{\varphi+1}}{(\varphi+1)(1-p)}$$

For the special case $p=0$, the formulas simplify considerably and we have:

$$\mu_\varphi = \frac{1}{\varphi+1} \quad \textit{and} \quad S_\varphi(t) = (1-t)^\varphi$$

$$\Rightarrow \mu_\varphi^{\,2} + \sigma_\varphi^{\,2} = 2\int_0^1 t(1-t)^\varphi dt = \frac{2\Gamma(2)\Gamma(\varphi+1)}{\Gamma(\varphi+3)} = \frac{2}{(\varphi+2)(\varphi+1)}$$

$$\Rightarrow \sigma_\varphi^{\,2} = \frac{\varphi}{(\varphi+1)^2(\varphi+2)}$$

In this case,

$$\sigma^2 = \frac{1}{12}; \qquad \sigma_\varphi = \sigma \Leftrightarrow \varphi \in \left\{1, \frac{\sqrt{33}-5}{2}\right\}$$

and the variance is maximized exactly when $\varphi = \frac{\sqrt{5}-1}{2}$, giving a specific illustration of the relationship between a proportional shift in the hazard and the variance.

The next example is a simple way to define a new hazard function from an old one.

Example IV.3. *Reversed hazard function,* let $h(t)$ be any hazard function of finite support and define $\bar{h}(t) = h(1-t)$, then clearly

$$\left(\bar{h}\right)_\varphi(t) = \varphi h(1-t) = \left(\overline{h_\varphi}\right)(t) \quad \text{for every } \varphi > 0$$

which shows that the reverse of a proportional hazard model is also one. Clearly, the reverse of an increasing (decreasing) hazard function is decreasing (increasing) and $\bar{\bar{h}} = h$. Letting $u = 1-t$, we find that

$$\bar{g}(t) = \int_0^t \bar{h}(s)ds = -\int_1^{1-t} h(u)du = g(1) - g(1-t)$$

$$\bar{p} = p; \quad \bar{S}(t) = \frac{p}{S(1-t)}; \quad \bar{\mu} = p\int_0^1 \frac{dt}{S(1-t)}$$

The reverse of Example IV.1 is, of course, again Example IV.1. The reverse of Example IV.2 is a decreasing hazard function with survival function $\frac{p}{1-(1-p)(1-t)}$ and mean $\frac{p\ln(p)}{p-1}$.

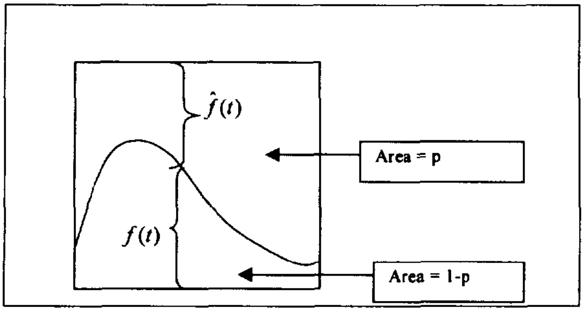
The next example is another simple way to define a new hazard function from an old one.

Example IV.4. *Complement hazard function,* let $h(t)$ be any hazard function of finite support such that $f(t) \le 1, 0 \le t \le 1$, and define $\hat{f}(t) = 1 - f(t)$, then

$$\hat{S}(t) = 1 - \int_0^t \hat{f}(s)ds = 1 - \int_0^t (1 - f(s))ds = 1 - t + (1 - S(t)) = 2 - t - S(t)$$

$$\hat{p} = 1 - p; \quad \hat{h}(t) = \frac{1 - f(t)}{2 - t - S(t)}$$

We again clearly have $\hat{\hat{f}}(t) = f(t)$; the picture is:



Example IV.5: *Modified Beta density with parameters* a, b, c, p. Assume $a, c > -1, b > 0, 0 \le p < 1$ and define

$$f(a,b,c,p;t) = \frac{b(1-p)t^a(1-t^b)^c}{\mathrm{B}\left(\frac{a+1}{b}, c+1\right)}$$

Then, clearly, $f(t) \ge 0$, when $0 \le t \le 1$ and the above lemma implies that

$$\int_0^1 f(a,b,c,p;t)dt = 1 - p$$

The binomial theorem enables us to write:

$$f(a,b,c,p;t)=\frac{b(1-p)}{\mathrm{B}\left(\frac{a+1}{b},c+1\right)}\sum_{k=0}^{\infty}(-1)^k\binom{c}{k}t^{a+bk}$$

$$S(a,b,c,p;t)=1-\int_0^t f(s)ds=1+\frac{b(1-p)}{\mathrm{B}\left(\frac{a+1}{b},c+1\right)}\sum_{k=0}^{\infty}\frac{(-1)^{k+1}\binom{c}{k}t^{a+bk+1}}{a+bk+1}$$

When c is an **integer**, the reduction formula for the Gamma function gives:

$$\mathrm{B}\left(\frac{a+1}{b},c+1\right)=\frac{\Gamma\left(\frac{a+1}{b}\right)\Gamma(c+1)}{\Gamma\left(\frac{a+1}{b}+c+1\right)}=\frac{\Gamma\left(\frac{a+1}{b}\right)c!}{\Gamma\left(\frac{a+1}{b}\right)\prod_{j=0}^{c}\left(\frac{a+1}{b}+j\right)}$$

$$=\frac{c!}{\prod_{j=0}^{c}\left(\frac{a+bj+1}{b}\right)}=\frac{b^{c+1}c!}{\prod_{j=0}^{c}a+bj+1}$$

from which we find that for c an **integer:**

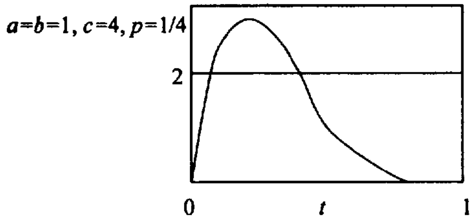
$$f(a,b,c,p;t)=\frac{b(1-p)\prod_{j=0}^{c}(a+bj+1)}{b^{c+1}c!}\sum_{k=0}^{c}(-1)^k\binom{c}{k}t^{a+bk}$$

$$=\frac{(1-p)\prod_{j=0}^{c}(a+bj+1)}{b^c}\sum_{k=0}^{c}\frac{(-1)^k t^{a+bk}}{k!(c-k)!}$$

$$S(a,b,c,p;t)=1+\frac{(1-p)}{b^c}\sum_{k=0}^{c}\prod_{\substack{j=0\\ j\neq k}}^{c}(a+bj+1)\frac{(-1)^{k+1}t^{a+bk+1}}{k!(c-k)!}$$

which expresses f(t) and S(t) as polynomials. When $ac\neq 0$, $f(t)=0\Leftrightarrow t\in\{0,1\}$. In fact, it is readily verified that $f(t)$ is positive on $(0,1)$ with a unique maximum at $t=\left(\frac{a}{a+cb}\right)^{\frac{1}{b}}$. It follows that the hazard function in this example is generally $\bigcap$-shaped. The following picture illustrates this:

h(t)



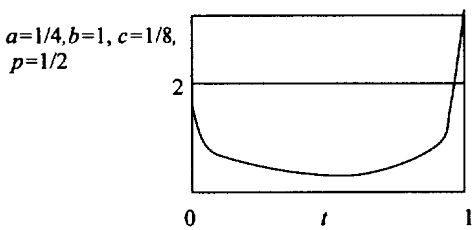
The final example is a slight variation of the previous one.

Example IV.6: Assume $a,c > -1, b > 0, 0 \le p < 1$ and define

$$f(a,b,c,p;t) = \frac{b(1-p)(1-t^a(1-t^b)^c)}{b - \mathrm{B}\left(\frac{a+1}{b}, c+1\right)}$$

Again $f(t) \ge 0$, when $0 \le t \le 1$ and the above lemma implies that $\int_0^1 f(a,b,c,p;t)dt = 1-p$. In this case, the hazard function is generally $\bigcup$-shaped. The following picture illustrates this:

h(t)



In the event that a particular shape of the hazard function is required, the last two examples provide candidates for parameter estimation. The following section argues that, for most purposes, a simple step function is preferable, from both the conceptual and computational perspectives.

This section concludes with two results. The first is one more observation on the difference $\mu - \mu_\delta$. The second revisits how for a finite hazard the survival function is a convenient device for computing moments, in this case relating it with the moment generating function.

In Example IV.2, note that

$$\mu-\mu_{\varphi}=\frac{p+1}{2}-\frac{p^{\varphi+1}-1}{(\varphi+1)(p-1)}=\frac{1}{1-p}\left[\frac{1-p^{2}}{2}+\frac{p^{\varphi+1}-1}{\varphi+1}\right]$$

The following generalizes this:

Proposition IV.3: Assume $f(t)$ is continuous on (0,1), then $\forall\varphi>0,\ \ 0\le\alpha<1,\ \ \exists\zeta\in(\alpha,1)$ such that

$$f(\zeta)(\mu-\mu_{\delta})=p_{\alpha}^{\ 2}\left[\frac{1-\left(\frac{p}{p_{\alpha}}\right)^{2}}{2}+\frac{\left(\frac{p}{p_{\alpha}}\right)^{\varphi+1}-1}{\varphi+1}\right]\quad \text{where } p_{\alpha}=S(\alpha).$$

Proof:

$$\mu-\mu_{\delta}=\int_{0}^{1}S(t)-S_{\delta}(t)dt=\int_{0}^{\alpha}S(t)-S(t)dt+\int_{\alpha}^{1}S(t)-S(\alpha)^{1-\varphi}S(t)^{\varphi}dt$$

$$=\int_{\alpha}^{1}S(t)-p_{\alpha}^{\ 1-\varphi}S(t)^{\varphi}dt$$

Consider first the case $\mu=\mu_{\delta}$. Observe that

$$S(t)-p_{\alpha}^{\ 1-\varphi}S(t)^{\varphi}\begin{cases}\le 0 & \varphi\le 1\\ \ge 0 & \varphi\ge 1\end{cases}$$

It follows, therefore, by continuity and the preceding equation, that $\mu=\mu_{\delta}$ would force

$$S(t)-p_{\alpha}^{\ 1-\varphi}S(t)^{\varphi}=0\quad \forall t\in(\alpha,1)$$

Now if $p=p_{\alpha}$, then the right hand side is clearly 0 and the result holds. So consider the case $\mu=\mu_{\delta},\ p_{\alpha}<p$. We then have both:

$$p-p_{\alpha}^{\ 1-\varphi}p^{\varphi}=\lim_{t\to 1}\left\{S(t)-p_{\alpha}^{\ 1-\varphi}S(t)^{\varphi}\right\}=\lim_{t\to 1}\{0\}=0$$

$$\text{and}\quad p-p_{\alpha}^{\ 1-\varphi}p^{\varphi}\begin{cases}<0 & \varphi<1\\ >0 & \varphi>1\end{cases}$$

which clearly forces $\varphi = 1$. The result again follows since $\varphi = 1$ makes the right hand side 0.

The upshot is that we may now assume that $\mu \neq \mu_\delta$. Because f is continuous and does not change sign, the generalized intermediate value theorem for integrals $\Rightarrow \exists \zeta \in (\alpha, 1)$ such that

$$f(\zeta)\int_\alpha^1 S(t) - p_\alpha^{1-\varphi} S(t)^\varphi dt = \int_\alpha^1 \left(S(t) - p_\alpha^{1-\varphi} S(t)^\varphi\right) f(t) dt, \quad f(\zeta) > 0$$

Noting that $dS = -f(t)dt;\quad t = \alpha \Leftrightarrow S(t) = p_\alpha;\quad t = 1 \Leftrightarrow S(t) = p$. With the change of variable we have:

$$f(\zeta)(\mu - \mu_\delta) = \int_\alpha^1 \left(S(t) - p_\alpha^{1-\varphi} S(t)^\varphi\right) f(t) dt = \int_p^{p_\alpha} \left(S - p_\alpha^{1-\varphi} S^\varphi\right) dS$$

$$= \left.\frac{S^2}{2} - p_\alpha^{1-\varphi} \frac{S^{\varphi+1}}{\varphi+1}\right]_p^{p_\alpha} = \frac{p_\alpha^2}{2} - p_\alpha^{1-\varphi} \frac{p_\alpha^{\varphi+1}}{\varphi+1} - \frac{p^2}{2} + p_\alpha^{1-\varphi} \frac{p^{\varphi+1}}{\varphi+1}$$

$$= p_\alpha^2 \left[\frac{1 - \left(\frac{p}{p_\alpha}\right)^2}{2} + \frac{\left(\frac{p}{p_\alpha}\right)^{\varphi+1} - 1}{\varphi+1} \right]$$

which completes the proof.

Proposition IV.4. $M_T(x) = 1 + x\int_0^1 e^{xt} S(t) dt$

Proof: By definition:

$$M_T(x) = E(e^{xT}) = \int_0^1 e^{xt} f(t) dt + pe^x$$

Under our assumptions, we can interchange summation with integration, whence:

$$
\begin{aligned}
M_T(x) - pe^x &= \int_0^1 e^{xt} f(t)dt \\
&= \int_0^1 \sum_{k=0}^{\infty} \frac{(xt)^k}{k!} f(t)dt \\
&= \sum_{k=0}^{\infty} \frac{x^k}{k!} \int_0^1 t^k f(t)dt \\
&= \int_0^1 f(t)dt + \sum_{k=1}^{\infty} \frac{x^k}{k!} \left[k\int_0^1 t^{k-1} S(t)dt - p \right] \\
&= 1 - p + x\sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} \left[\int_0^1 t^{k-1} S(t)dt - \frac{p}{k} \right] \\
&= 1 - p + x\sum_{k=1}^{\infty} \left[\int_0^1 \frac{x^{k-1}t^{k-1}}{(k-1)!} S(t)dt \right] - \sum_{k=1}^{\infty} \frac{px^k}{k!} \\
&= 1 + x\sum_{k=0}^{\infty} \left[\int_0^1 \frac{x^k t^k}{k!} S(t)dt \right] - p\sum_{k=0}^{\infty} \frac{x^k}{k!} \\
&= 1 + x\left[\int_0^1 \sum_{k=0}^{\infty} \frac{(xt)^k}{k!} S(t)dt \right] - pe^x \\
&= 1 + x\int_0^1 e^{xt} S(t)dt - pe^x .
\end{aligned}
$$

Section V: Combining Finite Support Hazard Functions

We continue with the notation and assumptions of the previous section. Consider first the case of two hazard functions, $h_1(t)$ and $h_1(t)$. If these represent independent causes of failure, then their sum $h_1 + h_2$ provides the corresponding hazard function. In this case, we clearly have:

$$g = g_1 + g_2; \quad S = S_1 S_2; \quad f = S_1 f_2 + f_1 S_2; \quad \mu = \int_0^1 S_1(t) S_2(t) dt,$$

and we can readily generalize this to the case of compounding together any finite number of hazards.

Consider the case of adding a constant hazard, i.e., the case $h_2(t) \equiv a > 0$. While this will clearly decrease the mean duration to failure, the issue is by how much. From Example IV.1, we have $S_2(t) = e^{-at}$, and from Proposition IV.4 we find:

$$M_{T_1}(-a) = 1 - a \int_0^1 e^{-at} S_1(t) dt = 1 - a \int_0^1 S_1(t) S_2(t) dt = 1 - a\mu \quad \Rightarrow \mu = \frac{1 - M_{T_1}(-a)}{a}$$

While adding hazards is formally very simple, this suggests that the effect of the mean duration can become complicated in even the simplest contexts. Moreover, the more useful and challenging task would be to reverse this process: to decompose a compound hazard into mutually independent hazards. Fortunately, our needs are much less demanding.

In this section we detail a very simple and straightforward way to combine hazard functions. This provides the framework needed to exploit the Cox Proportional hazard model to approximate hazard functions with step functions. The approach also fits in well within the context of time-dependent interventions.

Begin with a finite support hazard function $h(t)$ and let $\{0 = \alpha_0 < \alpha_1 < ... < \alpha_n = 1\}$ be a partition of $[0,1]$ into n subintervals. We can readily decompose $h(t)$ into n finite support hazard functions:

$$h_i(t) = h(\alpha_{i-1} + t(\alpha_i - \alpha_{i-1})) \quad 0 \le t \le 1, i = 1,2,...,n$$

Fortunately, this process is readily reversed, i.e. given an ordered set of n finite support hazard functions $\{h_i(t), i = 1,2,...,n\}$ together with a partition $\{0 = \alpha_0 < \alpha_1 < ... < \alpha_n = 1\}$ of $[0,1]$ into n subintervals, we define their *gauntlet hazard function* on $[0,1]$ by

$$h(t) = \{h_1, h_2, ..., h_n; 0 = \alpha_0 < \alpha_1 < ... < \alpha_n = 1\}(t) = h_i\left(\frac{t - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}}\right) \quad \text{where } \alpha_{i-1} \le t < \alpha_i$$

We observe that when the $h_i, i = 1,2,...,n$ are all constant hazard functions ($h_i(t) \equiv \varphi_i = h_{\varphi_i}, i = 1,2,...,n$ from Example IV.1) their gauntlet hazard function is a step function. Conversely, any hazard step function is the gauntlet of constant hazard functions in an essentially unique way.

The interpretation is straightforward. As suggested by the name, we can think of the hazards being lined up in sequence, much like a gauntlet. Survival becomes a matter of passing successively through the hazards, in sequence. A concern arises when any $p_i = 0, i < n$, since failure is assured during the corresponding interval, rendering the rest of the gauntlet essentially moot and introduces a singularity in the hazard function. As was noted before, the case $p=0$ is akin to infinite support hazards. In general, the probability of surviving the *i-th* interval of the gauntlet hazard is $\prod_{k=1}^{i} p_k$.

From these definitions, combined with our notational conventions, we have:

$$g(t)=\sum_{k=1}^{i-1}(\alpha_k-\alpha_{k-1})g_k(1)+(\alpha_i-\alpha_{i-1})g_i\left(\frac{t-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}\right) \quad \text{where } \alpha_{i-1}\le t<\alpha$$

$$S(t)=\prod_{k=1}^{i-1} p_k^{\alpha_k-\alpha_{k-1}}\, S_i\left(\frac{t-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}\right)^{\alpha_i-\alpha_{i-1}} \quad \text{where } \alpha_{i-1}\le t<\alpha$$

$$\mu=\int_0^1 S(t)dt=\sum_{i=1}^{n}\int_{\alpha_{i-1}}^{\alpha_i}S(t)dt=\sum_{i=1}^{n}\prod_{k=1}^{i-1}p_k^{\alpha_k-\alpha_{k-1}}\int_{\alpha_{i-1}}^{\alpha_i}S_i\left(\frac{t-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}\right)^{\alpha_i-\alpha_{i-1}}dt$$

$$=\sum_{i=1}^{n}\prod_{k=1}^{i-1}p_k^{\alpha_k-\alpha_{k-1}}(\alpha_i-\alpha_{i-1})\int_0^1 S_i(u)^{\alpha_i-\alpha_{i-1}}du$$

$$=\sum_{i=1}^{n}\prod_{k=1}^{i-1}p_k^{\alpha_k-\alpha_{k-1}}(\alpha_i-\alpha_{i-1})(\mu_i)_{\alpha_i-\alpha_{i-1}}$$

It is instructive to note two special cases of this formula for μ:

Case 1: Assume the partition is uniform, that is, $\alpha_i=\frac{i}{n}$ then the formula becomes:

$$\mu=\frac{1}{n}\sum_{i=1}^{n}\left(\prod_{k=1}^{i-1}p_k\right)^{\frac{1}{n}}(\mu_i)_{\frac{1}{n}}$$

Case 2: Assume the hazard is constant on all the intervals (step function). Then by Example IV.1,

$$h_i \equiv \varphi_i \Rightarrow (\mu_i)_{\alpha_i - \alpha_{i-1}} = \frac{1 - e^{\varphi_i(\alpha_{i-1} - \alpha_i)}}{\varphi_i(\alpha_i - \alpha_{i-1})},$$

and the formula becomes

$$\begin{aligned}
\mu &= \sum_{i=1}^{n} \prod_{k=1}^{i-1} p_k^{\alpha_k - \alpha_{k-1}} (\alpha_i - \alpha_{i-1})(\mu_i)_{\alpha_i - \alpha_{i-1}} \\
&= \sum_{i=1}^{n} \prod_{k=1}^{i-1} e^{\varphi_k(\alpha_{k-1} - \alpha_k)} (\alpha_i - \alpha_{i-1}) \left(\frac{1 - e^{\varphi_i(\alpha_{i-1} - \alpha_i)}}{\varphi_i(\alpha_i - \alpha_{i-1})} \right) \\
&= \sum_{i=1}^{n} e^{\sum_{k=1}^{i-1} \varphi_k(\alpha_{k-1} - \alpha_k)} \left(\frac{1 - e^{\varphi_i(\alpha_{i-1} - \alpha_i)}}{\varphi_i} \right)
\end{aligned}$$

Finally, when both apply, in the case of a step function with uniform partition, the formula simplifies to:

$$\mu = \sum_{i=1}^{n} e^{-\frac{1}{n}\sum_{k=1}^{i-1} \varphi_k} \left(\frac{1 - e^{-\frac{\varphi_i}{n}}}{\varphi_i} \right)$$

In the example below, we consider how to make use of this, given a set of empirical observations. The formulas suggest that it may prove useful to approximate the hazard function by a step function. In that regard, notice that the natural choice for $\varphi_i \approx h_i(t)$ is the average value of the hazard function over the ith interval. This, in turn, is readily determined from the survival function:

$$\frac{1}{\alpha_i - \alpha_{i-1}} \int_{\alpha_{i-1}}^{\alpha_i} h(t)\,dt = \frac{g(\alpha_i) - g(\alpha_{i-1})}{\alpha_i - \alpha_{i-1}} = \frac{\ln(S(\alpha_{i-1})) - \ln(S(\alpha_i))}{\alpha_i - \alpha_{i-1}}$$

We conclude with a simple example that illustrates how, despite the awkwardness of the formulas, the calculations can be quite simple in practice.

Example V.1 Let $h_0(t) \equiv 1$, for $0 \le t \le 1$ be the constant unit hazard function and $\delta = \delta(\frac{1}{3}, 2) \circ \delta(\frac{2}{3}, \frac{1}{2})$ be the composite of the two shifts. Consider the hazard step function defined from:

$$h(t) = (h_0)_s(t) = \begin{cases} 1 & 0 \le t < \frac{1}{3} \\ 2 & \frac{1}{3} \le t < \frac{2}{3} \\ 1 & \frac{2}{3} \le t \le 1 \end{cases}$$

SAS was used to simulate two survival data sets One and Two, conforming to the hazard functions h_0 and h, respectively. The PDFs are readily determined from earlier examples and were used to perform the simulations (refer to Appendix 1 for details). A survival function was produced from Two. An excerpt of the output is provided below (page 10 of the listing),

t	$S(t)$	$g(t)$
0	1	0
$1/3$	0.71665	$0.33317 \approx 1/3$
$2/3$	0.36770	$1.00048 \approx 1$
1	0.26359	$1.33335 \approx 1/3$

The estimation of the hazard function $h(t)$ from the survival function is:

$$t \in [0, 1/3], \quad h(t) \approx \frac{g(1/3) - g(0)}{1/3 - 0} \approx \frac{1/3 - 0}{1/3} = 1$$

$$t \in [1/3, 2/3], \quad h(t) \approx \frac{1 - 1/3}{1/3} = 2$$

$$t \in [2/3, 1], \quad h(t) \approx \frac{\frac{4}{3} - 1}{1/3} = 1$$

The simple average of an upper and a lower Riemann sum of the survival function over [0,1] (equivalent to the trapezoidal rules since the survival function is monotonically decreasing) was used to estimate the mean duration to failure to be 0.56193 (page 16 of the listing):

$$\mu = \int_0^1 S(t)dt \approx 0.56193$$

Compare this with the value determined using the above formula:

$\varphi_1 = \varphi_3 = 1, \varphi_2 = 2:$

$$\mu = \left(1 - e^{-\frac{1}{3}}\right) + e^{-\frac{1}{3}}\left(\frac{1 - e^{-\frac{2}{3}}}{2}\right) + e^{-\frac{1}{3}-\frac{2}{3}}\left(1 - e^{-\frac{1}{3}}\right)$$

$$= 0.562077$$

Finally, data set Two observations were flagged and pooled with set One survival data. The SAS PHREG procedure was then run on the combined data set with the flagged data modeled as a time-dependent intervention applicable to the middle interval. The PHREG procedure produced a hazard ratio of 2.000 (page 4 of the listing) for that intervention, illustrating how the Cox proportional hazard model can be used to approximate a hazard function by a step function. By the same token, it illustrates how that procedure may provide the means to unpack this process. More precisely, the procedure results may reveal a change in hazard as (approximated by) a combination of shifts like the ones considered here: $\delta = \delta(\frac{1}{3}, 2) \circ \delta(\frac{2}{3}, \frac{1}{2})$. From that, the results of this paper can be used to translate this into the effect on the mean time to failure.

References:

[1] Allison, Paul D., *Survival Analysis Using the SAS® System: A Practical Guide*, The SAS Institute, Inc., 1995.

[2] Hogg, Robert V. and Klugman, Stuart A., *Loss Distributions*, John Wiley & Sons, 1984.

[3] Wang, Shuan, "Implementation of Proportional Hazards Transforms in Ratemaking," PCAS LXXXV, pp. 940-979.

APPENDIX 1

SASLOG:

```
*******************************;
          OPTIONS MPRINT LS=131 PS=59 NOCENTER;
          *OPTIONS OBS = 100;
          DATA ZERO;
          INPUT Z;
          CARDS;

NOTE: The data set WORK.ZERO has 1 observations and 1 variables.

          ;
          DATA ONE;SET ZERO;
          KEEP T CLOSED SHOCK;
          RETAIN COUNT;
          IF _N_ = 1;
          CLOSED = 1;
          SHOCK  = 0;
          COUNT = 0;
          DO I = 1 TO 1000;
             T = I/1000;
             DO J = 1 TO ROUND(50*EXP(-T),1);
                COUNT + 1;OUTPUT;END;END;
          T = 1;P = EXP(-1);
          CLOSED = 0;
          DO J = 1 TO (P/(1-P))*COUNT;
             OUTPUT;END;

NOTE: The data set WORK.ONE has 49980 observations and 3 variables.

          DATA TWO;SET ZERO;
          KEEP T CLOSED SHOCK;
          RETAIN COUNT;
          IF _N_ = 1;
          CLOSED = 1;
          SHOCK  = 1;
          COUNT = 0;
          DO I = 1 TO 333;
             T = I/1000;
             DO J = 1 TO ROUND(50*EXP(-T),1);
                COUNT + 1;OUTPUT;END;END;
          DO I = 334 TO 666;
             T = I/1000;
             DO J = 1 TO ROUND(100*EXP(-2*T + 1/3),1);
                COUNT + 1;OUTPUT;END;END;
          DO I = 667 TO 1000;
             T = I/1000;
             DO J = 1 TO ROUND(50*EXP(-T - 1/3),1);
                COUNT + 1;OUTPUT;END;END;
          T = 1;P = EXP(-4/3);
          CLOSED = 0;
          DO J = 1 TO (P/(1-P))*COUNT;
             OUTPUT;END;

NOTE: The data set WORK.TWO has 49956 observations and 3 variables.
```

```
        DATA THREE; SET ONE TWO;
        TITLE 'PHREG PAPER:TEST';

NOTE: The data set WORK.THREE has 99936 observations and 3 variables.
        PROC PHREG SIMPLE DATA=THREE;
             MODEL T*CLOSED(0)= SHOCK /CORRB COVB;

NOTE: The PROCEDURE PHREG printed pages 1-2.

        PROC PHREG SIMPLE DATA=THREE;
             MODEL T*CLOSED(0)= TSHOCK /CORRB COVB;
             IF 2/3 >= T > 1/3 THEN TSHOCK = SHOCK;ELSE TSHOCK = 0;
        %MACRO MEANDUR;
        PROC PHREG SIMPLE DATA=ZDATA;
             MODEL T*CLOSED(0)= /CORRB COVB;
             BASELINE OUT=BASE SURVIVAL=S;
        DATA BASE;SET BASE END = EOF;
        IF _N_ = 1 THEN DO;T = 0;S = 1;OUTPUT;END;
        IF T < 1 THEN OUTPUT;
        IF EOF OR T >= 1 THEN DO;T = 1;OUTPUT;END;
        PROC SORT NODUP DATA = BASE; BY T;
        DATA SUBBASE;SET BASE;
        IF ABS(T -   0) < .01 OR
           ABS(T - 1/3) < .01 OR
           ABS(T - 2/3) < .01 OR
           ABS(T -   1) < .01;
        G = -LOG(S);
        PROC PRINT DATA=SUBBASE;
        DATA MEAN;SET BASE END=EOF;KEEP  UPPER LOWER MEAN;
        KEEP  UPPER LOWER MEAN;
        RETAIN UPPER LOWER OLD_S OLD_T;
        IF _N_ = 1 THEN DO;
           OLD_T  = 0;
           UPPER = 0;
           LOWER = 0;
           OLD_S = 1;
           END;
        D = T - OLD_T;
        IF D > 0 THEN DO;
           UPPER + D*OLD_S;
           LOWER + D*S;
           END;
           OLD_T = T;
           OLD_S = S;
        IF EOF THEN DO;
           MEAN = (UPPER + LOWER)/2;OUTPUT;
           END;
        PROC PRINT DATA = MEAN;
        %MEND MEANDUR;
        DATA ZDATA;SET ONE;
NOTE: The PROCEDURE PHREG printed pages 3-4.
NOTE: The PROCEDURE PHREG used 3001K.

        TITLE 'PHREG PAPER:TEST BASE ONE';
        %MEANDUR;

NOTE: The data set WORK.ZDATA has 49980 observations and 3 variables.
```

```
MPRINT(MEANDUR):    PROC PHREG SIMPLE DATA=ZDATA;
MPRINT(MEANDUR):    MODEL T*CLOSED(0)= /CORRB COVB;
MPRINT(MEANDUR):    BASELINE OUT=BASE SURVIVAL=S;
NOTE: There are no explanatory variables in the MODEL statement.
NOTE: The data set WORK.BASE has 1001 observations and 2 variables.
NOTE: The PROCEDURE PHREG printed page 5.
MPRINT(MEANDUR):    DATA BASE;
MPRINT(MEANDUR):   SET BASE END = EOF;
MPRINT(MEANDUR):    IF _N_ = 1 THEN DO;
MPRINT(MEANDUR):   T = 0;
MPRINT(MEANDUR):   S = 1;
MPRINT(MEANDUR):   OUTPUT;
MPRINT(MEANDUR):   END;
MPRINT(MEANDUR):    IF T < 1 THEN OUTPUT;
MPRINT(MEANDUR):    IF EOF OR T >= 1 THEN DO;
MPRINT(MEANDUR):   T = 1;
MPRINT(MEANDUR):   OUTPUT;
MPRINT(MEANDUR):   END;

NOTE: The data set WORK.BASE has 1002 observations and 2 variables.

MPRINT(MEANDUR):    PROC SORT NODUP DATA = BASE:
MPRINT(MEANDUR):    BY T;

NOTE: 1 duplicate observations were deleted.
NOTE: The data set WORK.BASE has 1001 observations and 2 variables.
MPRINT(MEANDUR):    DATA SUBBASE;
MPRINT(MEANDUR):   SET BASE;
MPRINT(MEANDUR):    IF ABS(T - 0) < .01 OR ABS(T - 1/3) < .01 OR ABS(T - 2/3) <
01 OR ABS(T - 1) < .01;
MPRINT(MEANDUR):    G = -LOG(S);

NOTE: The data set WORK.SUBBASE has 60 observations and 3 variables.
MPRINT(MEANDUR):    PROC PRINT DATA=SUBBASE;

NOTE: The PROCEDURE PRINT printed pages 6-7.
MPRINT(MEANDUR):    DATA MEAN;
MPRINT(MEANDUR):    SET BASE END=EOF;
MPRINT(MEANDUR):    KEEP UPPER LOWER MEAN;
MPRINT(MEANDUR):    RETAIN UPPER LOWER OLD_S OLD_T;
MPRINT(MEANDUR):    IF _N_ = 1 THEN DO;
MPRINT(MEANDUR):    OLD_T = 0;
MPRINT(MEANDUR):    UPPER = 0;
MPRINT(MEANDUR):    LOWER = 0;
MPRINT(MEANDUR):    OLD_S = 1;
MPRINT(MEANDUR):    END;
MPRINT(MEANDUR):    D = T - OLD_T;
MPRINT(MEANDUR):    IF D > 0 THEN DO;
MPRINT(MEANDUR):    UPPER + D*OLD_S;
MPRINT(MEANDUR):    LOWER + D*S;
MPRINT(MEANDUR):    END;
MPRINT(MEANDUR):    OLD_T = T;
MPRINT(MEANDUR):    OLD_S = S;
MPRINT(MEANDUR):    IF EOF THEN DO;
MPRINT(MEANDUR):    MEAN = (UPPER + LOWER)/2;
MPRINT(MEANDUR):    OUTPUT;
MPRINT(MEANDUR):    END;
```

```
NOTE: The data set WORK.MEAN has 1 observations and 3 variables.
MPRINT(MEANDUR):   PROC PRINT DATA = MEAN;

NOTE: The PROCEDURE PRINT printed page 8.
96          DATA ZDATA;SET TWO;
97          TITLE 'PHREG PAPER:TEST BASE TWO';
98          %MEANDUR;

NOTE: The data set WORK.ZDATA has 49956 observations and 3 variables.
MPRINT(MEANDUR):   PROC PHREG SIMPLE DATA=ZDATA;
MPRINT(MEANDUR):   MODEL T*CLOSED(0)= /CORRB COVB;
MPRINT(MEANDUR):   BASELINE OUT=BASE SURVIVAL=S;
NOTE: There are no explanatory variables in the MODEL statement.
NOTE: The data set WORK.BASE has 1001 observations and 2 variables.
NOTE: The PROCEDURE PHREG printed page 9.
MPRINT(MEANDUR):   DATA BASE;
MPRINT(MEANDUR):  SET BASE END = EOF;
MPRINT(MEANDUR):   IF _N_ = 1 THEN DO;
MPRINT(MEANDUR):  T = 0;
MPRINT(MEANDUR):  S = 1;
MPRINT(MEANDUR):  OUTPUT;
MPRINT(MEANDUR):  END;
MPRINT(MEANDUR):   IF T < 1 THEN OUTPUT;
MPRINT(MEANDUR):   IF EOF OR T >= 1 THEN DO;
MPRINT(MEANDUR):  T = 1;
MPRINT(MEANDUR):  OUTPUT;
MPRINT(MEANDUR):  END;

NOTE: The data set WORK.BASE has 1002 observations and 2 variables.
MPRINT(MEANDUR):   PROC SORT NODUP DATA = BASE;
MPRINT(MEANDUR):   BY T;

NOTE: HOST sort chosen, but SAS sort recommended.
NOTE: 1 duplicate observations were deleted.
NOTE: The data set WORK.BASE has 1001 observations and 2 variables.
MPRINT(MEANDUR):   DATA SUBBASE;
MPRINT(MEANDUR):  SET BASE;
MPRINT(MEANDUR):   IF ABS(T - 0) < .01 OR ABS(T - 1/3) < .01 OR ABS(T - 2/3) <
01 OR ABS(T - 1) < .01;
MPRINT(MEANDUR):   G = -LOG(S);

NOTE: The data set WORK.SUBBASE has 60 observations and 3 variables.
MPRINT(MEANDUR):   PROC PRINT DATA=SUBBASE;

NOTE: The PROCEDURE PRINT printed pages 10-11.
MPRINT(MEANDUR):   DATA MEAN;
MPRINT(MEANDUR):  SET BASE END=EOF;
MPRINT(MEANDUR):   KEEP UPPER LOWER MEAN;
MPRINT(MEANDUR):   RETAIN UPPER LOWER OLD_S OLD_T;
MPRINT(MEANDUR):   IF _N_ = 1 THEN DO;
MPRINT(MEANDUR):   OLD_T = 0;
MPRINT(MEANDUR):   UPPER = 0;
MPRINT(MEANDUR):   LOWER = 0;
MPRINT(MEANDUR):   OLD_S = 1;
MPRINT(MEANDUR):   END;
MPRINT(MEANDUR):   D = T - OLD_T;
```

```
MPRINT(MEANDUR):    IF D > 0 THEN DO;
MPRINT(MEANDUR):    UPPER + D*OLD_S;
MPRINT(MEANDUR):    LOWER + D*S;
MPRINT(MEANDUR):    END;
MPRINT(MEANDUR):    OLD_T = T;
MPRINT(MEANDUR):    OLD_S = S;
MPRINT(MEANDUR):    IF EOF THEN DO;
MPRINT(MEANDUR):    MEAN = (UPPER + LOWER)/2;
MPRINT(MEANDUR):    OUTPUT;
MPRINT(MEANDUR):    END;

NOTE: The data set WORK.MEAN has 1 observations and 3 variables.

MPRINT(MEANDUR):    PROC PRINT DATA = MEAN;

NOTE: The PROCEDURE PRINT printed page 12.

99          DATA ZDATA;SET THREE;
100         TITLE 'PHREG PAPER:TEST BASE THREE';
101         %MEANDUR;

NOTE: The data set WORK.ZDATA has 99936 observations and 3 variables.

MPRINT(MEANDUR):    PROC PHREG SIMPLE DATA=ZDATA;
MPRINT(MEANDUR):    MODEL T*CLOSED(0)= /CORRB COVB;
MPRINT(MEANDUR):    BASELINE OUT=BASE SURVIVAL=S;
NOTE: There are no explanatory variables in the MODEL statement.
NOTE: The data set WORK.BASE has 1001 observations and 2 variables.
NOTE: The PROCEDURE PHREG printed page 13.

MPRINT(MEANDUR):    DATA BASE;
MPRINT(MEANDUR):   SET BASE END = EOF;
MPRINT(MEANDUR):    IF _N_ = 1 THEN DO;
MPRINT(MEANDUR):   T = 0;
MPRINT(MEANDUR):   S = 1;
MPRINT(MEANDUR):   OUTPUT;
MPRINT(MEANDUR):   END;
MPRINT(MEANDUR):    IF T < 1 THEN OUTPUT;
MPRINT(MEANDUR):    IF EOF OR T >= 1 THEN DO;
MPRINT(MEANDUR):   T = 1;
MPRINT(MEANDUR):   OUTPUT;
MPRINT(MEANDUR):   END;

NOTE: The data set WORK.BASE has 1002 observations and 2 variables.

MPRINT(MEANDUR):    PROC SORT NODUP DATA = BASE;
MPRINT(MEANDUR):    BY T;

NOTE: 1 duplicate observations were deleted.
NOTE: The data set WORK.BASE has 1001 observations and 2 variables.

MPRINT(MEANDUR):    DATA SUBBASE;
MPRINT(MEANDUR):   SET BASE;
MPRINT(MEANDUR):    IF ABS(T - 0) < .01 OR ABS(T - 1/3) < .01 OR ABS(T - 2/3) <
01 OR ABS(T - 1) < .01;
MPRINT(MEANDUR):    G = -LOG(S);
```

```
NOTE: The data set WORK.SUBBASE has 60 observations and 3 variables.

MPRINT(MEANDUR):   PROC PRINT DATA=SUBBASE;

NOTE: The PROCEDURE PRINT printed pages 14-15.
MPRINT(MEANDUR):   DATA MEAN;
MPRINT(MEANDUR):  SET BASE END=EOF;
MPRINT(MEANDUR):  KEEP UPPER LOWER MEAN;
MPRINT(MEANDUR):   KEEP UPPER LOWER MEAN;
MPRINT(MEANDUR):   RETAIN UPPER LOWER OLD_S OLD_T;
MPRINT(MEANDUR):   IF _N_ = 1 THEN DO;
MPRINT(MEANDUR):   OLD_T = 0;
MPRINT(MEANDUR):   UPPER = 0;
MPRINT(MEANDUR):   LOWER = 0;
MPRINT(MEANDUR):   OLD_S = 1;
MPRINT(MEANDUR):   END;
MPRINT(MEANDUR):   D = T - OLD_T;
MPRINT(MEANDUR):   IF D > 0 THEN DO;
MPRINT(MEANDUR):   UPPER + D*OLD_S;
MPRINT(MEANDUR):   LOWER + D*S;
MPRINT(MEANDUR):   END;
MPRINT(MEANDUR):   OLD_T = T;
MPRINT(MEANDUR):   OLD_S = S;
MPRINT(MEANDUR):   IF EOF THEN DO;
MPRINT(MEANDUR):   MEAN = (UPPER + LOWER)/2;
MPRINT(MEANDUR):  OUTPUT;
MPRINT(MEANDUR):   END;

NOTE: The data set WORK.MEAN has 1 observations and 3 variables.

MPRINT(MEANDUR):   PROC PRINT DATA = MEAN;
NOTE: The PROCEDURE PRINT printed page 16.
```

SAS LISTING

PHREG PAPER:TEST page 1

The PHREG Procedure

Data Set: WORK.THREE
Dependent Variable: T
Censoring Variable: CLOSED
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
99936	68382	31554	31.57

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
SHOCK	99936	0.49988	0.50000	0	1.00000

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1510536.08	1509248.84	1287.233 with 1 DF (p=0.0001)
Score	.	.	1290.337 with 1 DF (p=0.0001)
Wald	.	.	1282.328 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
SHOCK	1	0.275302	0.00769	1282	0.0001	1.317

Estimated Covariance Matrix

	SHOCK
SHOCK	0.0000591043

Estimated Correlation Matrix

	SHOCK
SHOCK	1.000000000

The PHREG Procedure

Data Set: WORK.THREE
Dependent Variable: T
Censoring Variable: CLOSED
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
99936	68382	31554	31.57

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
TSHOCK	99936	0.17425	0.37933	0	1.00000

WARNING: Simple statistics listed for the time-dependent explanatory variables have limited value.

PHREG PAPER:TEST page 4

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1510536.08	1507331.64	3204.433 with 1 DF (p=0.0001)
Score	.	.	3197.243 with 1 DF (p=0.0001)
Wald	.	.	3073.086 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TSHOCK	1	0.693074	0.01250	3073	0.0001	2.000

Estimated Covariance Matrix

	TSHOCK
TSHOCK	0.0001563092

Estimated Correlation Matrix

	TSHOCK
TSHOCK	1.000000000

PHREG PAPER:TEST page 5

The PHREG Procedure

Data Set: WORK.ZDATA
Dependent Variable: T
Censoring Variable: CLOSED
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
49980	31594	18386	36.79

NOTE: There are no explanatory variables in this model.

-2 LOG L = 657271.7

PHREG PAPER: TEST BASE ONE pages 6-7

OBS	T	S	G
1	0.000	1.00000	0.00000
2	0.001	0.99900	0.00100
3	0.002	0.99800	0.00200
4	0.003	0.99700	0.00301
5	0.004	0.99600	0.00401
6	0.005	0.99500	0.00501
7	0.006	0.99400	0.00602
8	0.007	0.99300	0.00703
9	0.008	0.99200	0.00804
10	0.009	0.99100	0.00904
11	0.324	0.72327	0.32397
12	0.325	0.72255	0.32497
13	0.326	0.72183	0.32597
14	0.327	0.72111	0.32697
15	0.328	0.72039	0.32797
16	0.329	0.71967	0.32897
17	0.330	0.71895	0.32997
18	0.331	0.71823	0.33097
19	0.332	0.71751	0.33197
20	0.333	0.71679	0.33298
21	0.334	0.71607	0.33398
22	0.335	0.71535	0.33499
23	0.336	0.71463	0.33600
24	0.337	0.71391	0.33700
25	0.338	0.71319	0.33801
26	0.339	0.71247	0.33902
27	0.340	0.71174	0.34004
28	0.341	0.71102	0.34105
29	0.342	0.71030	0.34206
30	0.343	0.70960	0.34305
31	0.657	0.51845	0.65692
32	0.658	0.51793	0.65792
33	0.659	0.51741	0.65893
34	0.660	0.51689	0.65993
35	0.661	0.51637	0.66094
36	0.662	0.51585	0.66195
37	0.663	0.51533	0.66296
38	0.664	0.51481	0.66397
39	0.665	0.51429	0.66498
40	0.666	0.51377	0.66599
41	0.667	0.51325	0.66700
42	0.668	0.51273	0.66802
43	0.669	0.51220	0.66903
44	0.670	0.51168	0.67005
45	0.671	0.51116	0.67106
46	0.672	0.51064	0.67208
47	0.673	0.51012	0.67310
48	0.674	0.50962	0.67408
49	0.675	0.50912	0.67506
50	0.676	0.50862	0.67605
51	0.991	0.37117	0.99110
52	0.992	0.37079	0.99212
53	0.993	0.37041	0.99315
54	0.994	0.37003	0.99418
55	0.995	0.36967	0.99515
56	0.996	0.36931	0.99612
57	0.997	0.36895	0.99710
58	0.998	0.36859	0.99808
59	0.999	0.36823	0.99905
60	1.000	0.36787	1.00003

PHREG PAPER: TEST BASE ONE page 8

OBS	UPPER	LOWER	MEAN
1	0.63244	0.63180	0.63212

PHREG PAPER: TEST BASE TWO page 9

The PHREG Procedure

Data Set: WORK.ZDATA
Dependent Variable: T
Censoring Variable: CLOSED
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
49956	36788	13168	26.36

NOTE: There are no explanatory variables in this model.

-2 LOG L = 757604.6

OBS	T	S	G
1	0.000	1.00000	0.00000
2	0.001	0.99900	0.00100
3	0.002	0.99800	0.00200
4	0.003	0.99700	0.00301
5	0.004	0.99600	0.00401
6	0.005	0.99500	0.00502
7	0.006	0.99399	0.00602
8	0.007	0.99299	0.00703
9	0.008	0.99199	0.00804
10	0.009	0.99099	0.00905
11	0.324	0.72314	0.32416
12	0.325	0.72242	0.32515
13	0.326	0.72170	0.32615
14	0.327	0.72097	0.32715
15	0.328	0.72025	0.32815
16	0.329	0.71953	0.32915
17	0.330	0.71881	0.33015
18	0.331	0.71809	0.33116
19	0.332	0.71737	0.33216
20	0.333	0.71665	0.33317
21	0.334	0.71521	0.33518
22	0.335	0.71379	0.33717
23	0.336	0.71237	0.33916
24	0.337	0.71095	0.34116
25	0.338	0.70952	0.34316
26	0.339	0.70810	0.34517
27	0.340	0.70668	0.34717
28	0.341	0.70526	0.34919
29	0.342	0.70386	0.35118
30	0.343	0.70246	0.35317
31	0.657	0.37473	0.98155
32	0.658	0.37399	0.98353
33	0.659	0.37325	0.98551
34	0.660	0.37251	0.98750
35	0.661	0.37177	0.98949
36	0.662	0.37103	0.99148
37	0.663	0.37029	0.99348
38	0.664	0.36955	0.99548
39	0.665	0.36880	0.99749
40	0.666	0.36806	0.99950
41	0.667	0.36770	1.00048
42	0.668	0.36734	1.00146
43	0.669	0.36698	1.00244
44	0.670	0.36662	1.00342
45	0.671	0.36626	1.00441
46	0.672	0.36590	1.00539
47	0.673	0.36554	1.00637
48	0.674	0.36518	1.00736
49	0.675	0.36482	1.00835
50	0.676	0.36446	1.00934
51	0.991	0.26593	1.32451
52	0.992	0.26567	1.32549
53	0.993	0.26541	1.32647
54	0.994	0.26515	1.32745
55	0.995	0.26489	1.32843
56	0.996	0.26463	1.32941
57	0.997	0.26437	1.33040
58	0.998	0.26411	1.33138
59	0.999	0.26385	1.33237
60	1.000	0.26359	1.33335

PHREG PAPER: TEST BASE TWO **page 12**

OBS	UPPER	LOWER	MEAN
1	0.56229	0.56156	0.56193

PHREG PAPER: TEST BASE THREE page 13

The PHREG Procedure

Data Set: WORK.ZDATA
Dependent Variable: T
Censoring Variable: CLOSED
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
99936	68382	31554	31.57

NOTE: There are no **explanatory** variables **in this model.**

-2 LOG L = 1510536

PHREG PAPER: TEST BASE THREE **pages 14-15**

OBS	T	S	G
1	0.000	1.00000	0.00000
2	0.001	0.99900	0.00100
3	0.002	0.99800	0.00200
4	0.003	0.99700	0.00301
5	0.004	0.99600	0.00401
6	0.005	0.99500	0.00502
7	0.006	0.99400	0.00602
8	0.007	0.99300	0.00703
9	0.008	0.99199	0.00804
10	0.009	0.99099	0.00905
11	0.324	0.72320	0.32407
12	0.325	0.72248	0.32506
13	0.326	0.72176	0.32606
14	0.327	0.72104	0.32706
15	0.328	0.72032	0.32806
16	0.329	0.71960	0.32906
17	0.330	0.71888	0.33006
18	0.331	0.71816	0.33106
19	0.332	0.71744	0.33207
20	0.333	0.71672	0.33307
21	0.334	0.71564	0.33458
22	0.335	0.71457	0.33608
23	0.336	0.71350	0.33758
24	0.337	0.71243	0.33908
25	0.338	0.71136	0.34058
26	0.339	0.71028	0.34209
27	0.340	0.70921	0.34360
28	0.341	0.70814	0.34511
29	0.342	0.70708	0.34661
30	0.343	0.70603	0.34809
31	0.657	0.44661	0.80608
32	0.658	0.44598	0.80749
33	0.659	0.44535	0.80891
34	0.660	0.44471	0.81032
35	0.661	0.44408	0.81174
36	0.662	0.44345	0.81316
37	0.663	0.44282	0.81458
38	0.664	0.44219	0.81601
39	0.665	0.44156	0.81744
40	0.666	0.44093	0.81886
41	0.667	0.44049	0.81986
42	0.668	0.44005	0.82086
43	0.669	0.43961	0.82186
44	0.670	0.43917	0.82287
45	0.671	0.43873	0.82387
46	0.672	0.43829	0.82487
47	0.673	0.43785	0.82588
48	0.674	0.43742	0.82686
49	0.675	0.43699	0.82785
50	0.676	0.43656	0.82883
51	0.991	0.31856	1.14393
52	0.992	0.31824	1.14494
53	0.993	0.31792	1.14594
54	0.994	0.31760	1.14695
55	0.995	0.31729	1.14793
56	0.996	0.31698	1.14891
57	0.997	0.31667	1.14989
58	0.998	0.31636	1.15087
59	0.999	0.31605	1.15185
60	1.000	0.31574	1.15283

PHREG PAPER: TEST BASE THREE **page 16**

OBS	UPPER	LOWER	MEAN
1	0.59737	0.59669	0.59703

APPENDIX 2

SASLOG:

```
350        ********************************************************************;
351        ***BEGIN CODE FOR  CASE STUDY SECTION****************************;
352        %MACRO VLIST;
353               EMPL2
354               AY93-AY94
355               MF01 EC01
356               NOI_SPR NOI_CUT
357        %MEND VLIST;

NOTE: The data set WORK.ONE has 12512 observations and 100 variables.
358        DATA ONE;SET ONE;
359        KEEP T CLOSED01 TPA LAG2TPA %VLIST ;
MPRINT(VLIST):   EMPL2 AY93-AY94 MF01 EC01 NOI_SPR NOI_CUT
360        *CREATE BASELINE SURVIVAL FUNCTION FOR VANISHING COVARIATES';
361        *TPA IS NON TIME-DEPENDENT REFERRAL VARIABLE';
362        TITLE 'PROPORTIONAL HAZARD MODEL FOR BASELINE';

NOTE: The data set WORK.ONE has 12512 observations and 11 variables.
363        PROC SORT DATA=ONE;BY TPA;

NOTE: The data set WORK.ONE has 12512 observations and 11 variables.

364        DATA INRISK;
365        INPUT %VLIST TPA;
MPRINT(VLIST):   EMPL2 AY93-AY94 MF01 EC01 NOI_SPR NOI_CUT
366        CARDS;

NOTE: The data set WORK.INRISK has 1 observations and 8 variables.
368        ;
369        PROC PHREG SIMPLE DATA=ONE;
370             MODEL T*CLOSED01(0)= %VLIST TPA;
MPRINT(VLIST):   EMPL2 AY93-AY94 MF01 EC01 NOI_SPR NOI_CUT
371             BASELINE COVARIATES=INRISK OUT=BASE SURVIVAL=S / NOMEAN;
NOTE: The data set WORK.BASE has 958 observations and 10 variables.
NOTE: The PROCEDURE PHREG printed pages 1-2.

372        DATA BASE;SET BASE;KEEP  T  S;IF T > 0;
NOTE: The data set WORK.BASE has 957 observations and 2 variables.
 373        PROC SORT DATA = BASE; BY T;

NOTE: The data set WORK.BASE has 957 observations and 2 variables.
374        DATA BASE;SET BASE END = EOF;
375        IF _N_ = 1 THEN DO;T = 0;S = 1;OUTPUT;END;
376        IF T < 1 THEN OUTPUT;
377        IF EOF OR T >= 1 THEN DO;T = 1;OUTPUT;END;

NOTE: The data set WORK.BASE has 959 observations and 2 variables.
378        PROC SORT NODUP DATA = BASE; BY T;
379        *CAPTURE BASELINE SURVIVAL FUNCTION ON [0,1] TO ARRAY TABLE';


NOTE: The data set WORK.BASE has 958 observations and 2 variables.
```

```
        DATA BASE;SET BASE END=EOF;
        ARRAY MATT(I) T1-T1000;
        ARRAY MATS(I)` S1-S1000;
        KEEP   T1-T1000 S1-S1000;
        RETAIN T1-T1000 S1-S1000;
        I = MIN(_N_,1000);
        MATT = T;
        MATS = S;
        IF EOF THEN DO;
           DO I = _N_+ 1 TO 1000;
              MATT = 1;
              MATS = 0;
              END;
           OUTPUT;
           END;
        *RUN PROPOTIONAL HAZARD MODEL';
        *TXPA X=1,2,3 ARE TIME-DEPENDENT REFERRAL VARIABLES';
        TITLE 'PROPORTIONAL HAZARD MODEL WITH TIME DEPENDENT REFERRAL';

NOTE: The data set WORK.BASE has 1 observations and 2000 variables.
        PROC PHREG SIMPLE DATA=ONE OUTEST=PARMS;
             MODEL T*CLOSED01(0)= %VLIST
MPRINT(VLIST):   EMPL2 AY93-AY94 MF01 EC01 NOI_SPR NOI_CUT
                                      T1TPA T2TPA T3TPA;
             IF TPA=1 & T >= LAG2TPA THEN TTPA=1;ELSE TTPA = 0;
             IF 1/6  > T        THEN T1TPA=TTPA;ELSE T1TPA = 0;
             IF 1/3  > T >= 1/6 THEN T2TPA=TTPA;ELSE T2TPA = 0;
             IF        T >= 1/3 THEN T3TPA=TTPA;ELSE T3TPA = 0;
        *DETERMINE PHI=REFERRAL RISK RATIO BY TLAYER;
        TITLE 'HAZARD RATIO PHI BY TIME LAYER';
        DATA PARMS;SET PARMS;
NOTE: The data set WORK.PARMS has 1 observations and 14 variables.
NOTE: The PROCEDURE PHREG printed pages 3-4.

        KEEP TLAYER PHI;
        TLAYER = 1;PHI = EXP(T1TPA);OUTPUT;
        TLAYER = 2;PHI = EXP(T2TPA);OUTPUT;
        TLAYER = 3;PHI = EXP(T3TPA);OUTPUT;

NOTE: The data set WORK.PARMS has 3 observations and 2 variables.
        PROC PRINT DATA = PARMS;

NOTE: The PROCEDURE PRINT printed page 5.
        DATA ONE;SET ONE;
             IF 1/6  > T        THEN TLAYER = 1;
                                ELSE
             IF 1/3  > T >= 1/6 THEN TLAYER = 2;
                                ELSE
                                     TLAYER = 3;

NOTE: The data set WORK.ONE has 12512 observations and 12 variables.
        PROC SORT DATA=ONE  ;BY TLAYER;


NOTE: The data set WORK.ONE has 12512 observations and 12 variables.
        PROC SORT DATA=PARMS;BY TLAYER;
```

```
NOTE: The data set WORK.PARMS has 3 observations and 2 variables.
421         DATA ONE;MERGE ONE(IN=INO) PARMS(IN=INP);BY TLAYER;
422         IF INO & INP;

NOTE: The data set WORK.ONE has 12512 observations and 13 variables.
423         PROC SORT DATA=ONE; BY TPA;
424         *USE PHI AND BASELINE SURVIVAL ARRAY TO ADJUST T;

NOTE: The data set WORK.ONE has 12512 observations and 13 variables.
425         DATA ONE;SET ONE;
426         RETAIN T1-T1000 S1-S1000;
427         ARRAY MATT(I) T1 - T1000;
428         ARRAY MATS(I) S1 - S1000;
429         DROP T1-T1000 S1-S1000;
430         IF _N_ =1 THEN SET BASE;
431         IF (TPA = 1) & (T < 1) THEN DO;
432            ALPHA = LAG2TPA;
433            LOOKUP = 0;I = 1;
434            DO WHILE(LOOKUP = 0);
435               LHT = MATT;LHS = MATS;
436               I + 1;RHT = MATT;RHS = MATS;
437               IF LHT <= ALPHA <= RHT THEN DO;
438                  S_ALPHA = LHS + ((ALPHA - LHT)/(RHT - LHT))*(RHS - LHS);
439                  LOOKUP = 1;
440               END;
441            END;
442            LOOKUP = 0;I = 1;
443            DO WHILE(LOOKUP = 0);
444               LHT = MATT;LHS = MATS;
445               I + 1;RHT = MATT;RHS = MATS;
446               IF LHT <= T <= RHT THEN DO;
447                  S_T = LHS + ((T - LHT)/(RHT - LHT))*(RHS - LHS);
448                  LOOKUP = 1;
449               END;
450            END;
451            S_ADJT = (S_T**PHI)*(S_ALPHA**(1-PHI));
452            LOOKUP = 0;I = 1;
453            DO WHILE(LOOKUP = 0);
454               LHT = MATT;LHS = MATS;
455               I + 1;RHT = MATT;RHS = MATS;
456               IF LHS >= S_ADJT >= RHS THEN DO;
457                  ADJT = RHT + ((S_ADJT - RHS)/(LHS - RHS))*(LHT - RHT);
458                  LOOKUP = 1;
459               END;
460            END;
461         END;
462         ELSE DO;
463            ADJT = T;
464         END;
465         *USE PHREG TO MAKE SURVIVAL FUNCTION AT MEANS FOR ACTUAL DURATION T;
466         TITLE 'ACTUAL MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS';

NOTE: The data set WORK.ONE has 12512 observations and 24 variables.
467         PROC PHREG SIMPLE DATA=ONE;BY TPA;
468              MODEL T*CLOSED01(0)= %VLIST;
MPRINT(VLIST):   EMPL2 AY93-AY94 MF01 EC01 NOI_SPR NOI_CUT
469              BASELINE OUT=BASE SURVIVAL=S;
```

```
NOTE: The data set WORK.BASE has 1325 observations and 10 variables.
NOTE: The PROCEDURE PHREG printed pages 6-9.

          DATA MEAN;SET BASE;BY TPA;KEEP  TPA UPPER LOWER MEAN;
          RETAIN UPPER LOWER OLD_T OLD_S;
          IF FIRST.TPA THEN DO;
             UPPER = 0;
             LOWER = 0;
             OLD_T = 0;
             OLD_S = 1;
             END;
          IF OLD_T < T THEN DO;
             UPPER + (T - OLD_T)*OLD_S;
             LOWER + (T - OLD_T)*S;
             END;
             OLD_T = T;
             OLD_S = S;
          IF LAST.TPA THEN DO;
             MEAN = (UPPER + LOWER)/2;OUTPUT;
             END;

NOTE: The data set WORK.MEAN has 2 observations and 4 variables.

          DATA MEAN;SET MEAN;*CONVERT TO YEARS;
          UPPER = 3*UPPER;
          LOWER = 3*LOWER;
          MEAN  = 3*MEAN;

NOTE: The data set WORK.MEAN has 2 observations and 4 variables.
          PROC PRINT DATA = MEAN;
          *USE PHREG TO MAKE SURVIVAL FUNCTION AT MEANS FOR ADJUSTED DURATION;

NOTE: The PROCEDURE PRINT printed page 10.
          PROC PHREG SIMPLE DATA=ONE;BY TPA;
               MODEL ADJT*CLOSED01(0)= %VLIST;
MPRINT(VLIST):   EMPL2 AY93-AY94 MF01 EC01 NOI_SPR NOI_CUT
          BASELINE OUT=BASE SURVIVAL=S;
          TITLE 'ADJUSTED MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS';

NOTE: The data set WORK.BASE has 2311 observations and 10 variables.
NOTE: The PROCEDURE PHREG printed pages 11-14.

          DATA BASE;
          SET BASE;*RENAME ADJT TO BE SIMPLY T AS ABOVE;
          T = ADJT;

NOTE: The data set WORK.BASE has 2311 observations and 11 variables.
          DATA MEAN;SET BASE;BY TPA;KEEP  TPA UPPER LOWER MEAN;
          RETAIN UPPER LOWER OLD_T OLD_S;
          IF FIRST.TPA THEN DO;
             UPPER = 0;
             LOWER = 0;
             OLD_T = 0;
             OLD_S = 1;
             END;
          IF OLD_T < T THEN DO;
             UPPER + (T - OLD_T)*OLD_S;
```

```
            LOWER + (T - OLD_T)*S;
            END;
            OLD_T = T;
            OLD_S = S;
         IF LAST.TPA THEN DO;
            MEAN = (UPPER + LOWER)/2;OUTPUT;
            END;

NOTE: The data set WORK.MEAN has 2 observations and 4 variables.
         DATA MEAN;SET MEAN;*CONVERT TO YEARS;
         UPPER = 3*UPPER;
         LOWER = 3*LOWER;
         MEAN  = 3*MEAN;

NOTE: The data set WORK.MEAN has 2 observations and 4 variables.
         PROC PRINT DATA = MEAN;
NOTE: The PROCEDURE PRINT printed page 15.
```

SAS LISTING:

```
PROPORTIONAL HAZARD MODEL FOR BASELINE      page 1
 The PHREG Procedure

 Data Set: WORK.ONE
 Dependent Variable: T
 Censoring Variable: CLOSED01
 Censoring Value(s): 0
 Ties Handling: BRESLOW
```

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
12512	9961	2551	20.39

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
EMPL2	12512	0.10526	0.30690	0	1.00000
AY93	12512	0.33840	0.47318	0	1.00000
AY94	12512	0.33048	0.47041	0	1.00000
MF01	12512	0.57297	0.49467	0	1.00000
EC01	12512	0.56138	0.49624	0	1.00000
NOI_SPR	12512	0.68782	0.46340	0	1.00000
NOI_CUT	12512	0.24680	0.43117	0	1.00000
TPA	12512	0.17927	0.38359	0	1.00000

PROPORTIONAL HAZARD MODEL FOR BASELINE page 2
The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	171122.136	170068.900	1053.235 with 8 DF (p=0.0001)
Score	.	.	1052.856 with 8 DF (p=0.0001)
Wald	.	.	1034.622 with 8 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
EMPL2	1	-0.275576	0.03598	58.67093	0.0001	0.759
AY93	1	0.347298	0.02421	205.72217	0.0001	1.415
AY94	1	0.591461	0.03204	340.80216	0.0001	1.807
MF01	1	-0.208174	0.02216	88.26442	0.0001	0.812
EC01	1	0.245834	0.02227	121.89395	0.0001	1.279
NOI_SPR	1	0.140682	0.04307	10.66704	0.0011	1.151
NOI_CUT	1	0.186939	0.04580	16.65742	0.0001	1.206
TPA	1	0.134552	0.03398	15.68328	0.0001	1.144

The PHREG Procedure

Data Set: WORK.ONE
Dependent Variable: T
Censoring Variable: CLOSED01
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
12512	9961	2551	20.39

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
EMPL2	12512	0.10526	0.30690	0	1.00000
AY93	12512	0.33840	0.47318	0	1.00000
AY94	12512	0.33048	0.47041	0	1.00000
MF01	12512	0.57297	0.49467	0	1.00000
EC01	12512	0.56138	0.49624	0	1.00000
NOI_SPR	12512	0.68782	0.46340	0	1.00000
NOI_CUT	12512	0.24680	0.43117	0	1.00000
T1TPA	12512	0.09407	0.29194	0	1.00000
T2TPA	12512	0.06554	0.24748	0	1.00000
T3TPA	12512	0.01966	0.13884	0	1.00000

WARNING: Simple statistics listed for the time-dependent explanatory variables have limited value.

HAZARD RATIO PHI BY TIME LAYER page 4

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	171122.136	170012.885	1109.251 with 10 DF (p=0.0001)
Score	.	.	1125.216 with 10 DF (p=0.0001)
Wald	.	.	1101.663 with 10 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
EMPL2	1	-0.276409	0.03598	59.00264	0.0001	0.759
AY93	1	0.343170	0.02424	200.35078	0.0001	1.409
AY94	1	0.535551	0.03216	277.34071	0.0001	1.708
MF01	1	-0.207791	0.02216	87.95253	0.0001	0.812
EC01	1	0.245800	0.02227	121.79944	0.0001	1.279
NOI_SPR	1	0.143863	0.04308	11.15294	0.0008	1.155
NOI_CUT	1	0.189546	0.04581	17.11765	0.0001	1.209
T1TPA	1	0.353487	0.04301	67.54561	0.0001	1.424
T2TPA	1	0.184415	0.05265	12.26748	0.0005	1.203
T3TPA	1	0.114674	0.10755	1.13683	0.2863	1.122

HAZARD RATIO PHI BY TIME LAYER page 5

OBS	TLAYER	PHI
1	1	1.42402
2	2	1.20251
3	3	1.12151

TPA=0

The PHREG Procedure

Data Set: WORK.ONE
Dependent Variable: T
Censoring Variable: CLOSED01
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
10269	8544	1725	16.80

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
EMPL2	10269	0.11014	0.31308	0	1.00000
AY93	10269	0.38631	0.48693	0	1.00000
AY94	10269	0.21180	0.40861	0	1.00000
MF01	10269	0.57162	0.49487	0	1.00000
EC01	10269	0.56062	0.49634	0	1.00000
NOI_SPR	10269	0.68410	0.46490	0	1.00000
NOI_CUT	10269	0.25280	0.43464	0	1.00000

ACTUAL MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 7

TPA=0

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	143854.615	143106.464	748.151 with 7 DF (p=0.0001)
Score	.	.	743.549 with 7 DF (p=0.0001)
Wald	.	.	733.247 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
EMPL2	1	-0.289695	0.03776	58.86290	0.0001	0.748
AY93	1	0.341809	0.02451	194.49158	0.0001	1.407
AY94	1	0.573640	0.03373	289.16531	0.0001	1.775
MF01	1	-0.195073	0.02391	66.55687	0.0001	0.823
EC01	1	0.230103	0.02395	92.26850	0.0001	1.259
NOI_SPR	1	0.149435	0.04673	10.22562	0.0014	1.161
NOI_CUT	1	0.211587	0.04956	18.22332	0.0001	1.236

ACTUAL MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 8

TPA=1

The PHREG Procedure

Data Set: WORK.ONE
Dependent Variable: T
Censoring Variable: CLOSED01
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
2243	1417	826	36.83

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
EMPL2	2243	0.08292	0.27583	0	1.00000
AY93	2243	0.11904	0.32390	0	1.00000
AY94	2243	0.87383	0.33212	0	1.00000
MF01	2243	0.57914	0.49381	0	1.00000
EC01	2243	0.56487	0.49588	0	1.00000
NOI_SPR	2243	0.70486	0.45621	0	1.00000
NOI_CUT	2243	0.21935	0.41390	0	1.00000

ACTUAL MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 9

TPA=1

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	19580.571	19429.633	150.938 with 7 DF (p=0.0001)
Score	.	.	138.667 with 7 DF (p=0.0001)
Wald	.	.	131.971 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
EMPL2	1	-0.033574	0.11954	0.07888	0.7788	0.967
AY93	1	1.451228	0.38774	14.00880	0.0002	4.268
AY94	1	1.734945	0.38948	19.84265	0.0001	5.669
MF01	1	-0.281972	0.05898	22.85722	0.0001	0.754
EC01	1	0.342128	0.06101	31.44825	0.0001	1.408
NOI_SPR	1	0.091231	0.11133	0.67149	0.4125	1.096
NOI_CUT	1	0.033698	0.12062	0.07805	0.7800	1.034

ACTUAL MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 10

OBS	TPA	UPPER	LOWER	MEAN
1	0	1.02820	1.02542	1.02681
2	1	0.74414	0.73050	0.73732

ADJUSTED MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 11
TPA=0

The PHREG Procedure

Data Set: WORK.ONE
Dependent Variable: ADJT
Censoring Variable: CLOSED01
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
10269	8544	1725	16.80

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
EMPL2	10269	0.11014	0.31308	0	1.00000
AY93	10269	0.38631	0.48693	0	1.00000
AY94	10269	0.21180	0.40861	0	1.00000
MF01	10269	0.57162	0.49487	0	1.00000
EC01	10269	0.56062	0.49634	0	1.00000
NOI_SPR	10269	0.68410	0.46490	0	1.00000
NOI_CUT	10269	0.25280	0.43464	0	1.00000

TPA=0

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	143854.615	143106.464	748.151 with 7 DF (p=0.0001)
Score	.	.	743.549 with 7 DF (p=0.0001)
Wald	.	.	733.247 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
EMPL2	1	-0.289695	0.03776	58.86290	0.0001	0.748
AY93	1	0.341809	0.02451	194.49158	0.0001	1.407
AY94	1	0.573640	0.03373	289.16531	0.0001	1.775
MF01	1	-0.195073	0.02391	66.55687	0.0001	0.823
EC01	1	0.230103	0.02395	92.26850	0.0001	1.259
NOI_SPR	1	0.149435	0.04673	10.22562	0.0014	1.161
NOI_CUT	1	0.211587	0.04956	18.22332	0.0001	1.236

ADJUSTED MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 13

TPA=1

The PHREG Procedure

Data Set: WORK.ONE
Dependent Variable: ADJT
Censoring Variable: CLOSED01
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
2243	1417	826	36.83

Simple Statistics for Explanatory Variables

Total Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
EMPL2	2243	0.08292	0.27583	0	1.00000
AY93	2243	0.11904	0.32390	0	1.00000
AY94	2243	0.87383	0.33212	0	1.00000
MF01	2243	0.57914	0.49381	0	1.00000
EC01	2243	0.56487	0.49588	0	1.00000
NOI_SPR	2243	0.70486	0.45621	0	1.00000
NOI_CUT	2243	0.21935	0.41390	0	1.00000

ADJUSTED MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 14

TPA=1

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	19564.220	19414.420	149.800 with 7 DF (p=0.0001)
Score	.	.	137.480 with 7 DF (p=0.0001)
Wald	.	.	130.785 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
EMPL2	1	-0.027966	0.11955	0.05472	0.8150	0.972
AY93	1	1.456546	0.38758	14.12275	0.0002	4.291
AY94	1	1.738075	0.38932	19.93071	0.0001	5.686
MF01	1	-0.281598	0.05895	22.81847	0.0001	0.755
EC01	1	0.339608	0.06097	31.02579	0.0001	1.404
NOI_SPR	1	0.089939	0.11134	0.65252	0.4192	1.094
NOI_CUT	1	0.033703	0.12061	0.07808	0.7799	1.034

ADJUSTED MEAN DURATION FROM SURVIVAL FUNCTION AT MEANS page 15

OBS	TPA	UPPER	LOWER	MEAN
1	0	1.02820	1.02542	1.02681
2	1	0.83222	0.81883	0.82552

Modeling Multi-Dimensional Survival with Hazard Vector Fields

Daniel R. Corro

Abstract:

Traditional survival analysis deals with functions of one variable, "time." This paper explains the case of multiple and interacting aging metrics by introducing the notion of a hazard vector field. This approach is shown to provide a more general framework than traditional survival analysis, including the ability to model multi-dimensional censored data. A simple example illustrates how Green's Theorem in the plane applies to evaluate and even to theoretically optimize a course of action. One evident application is to the evaluation and promulgation of claim administration protocols.

Keywords: survival, vector field, hazard, gradient, line integral

Introduction

Although survival analysis has long recognized the need to account for different causes of death or failure, it recognizes only one way of measuring age. Consequently, survival functions—even "select" survival functions—are functions of one variable, typically denoted "t" and interpreted as "time". This paper explains the need to study observed lives from multiple perspectives. For example, a vehicle may burn several different types of fuel with varying and inter-related consumption patterns. The ability to determine whether a particular trip is possible and if so to find an efficient route may be best approached as a multi-dimensional problem. That is, it may not always be practical or revealing to reduce survival into functions of a single variable.

This work evolved from studying workers compensation insurance claims data and the motivation comes from that context. A quick claim resolution may not achieve a cost-effective result for either the insurer or the injured worker. A useful measure of "age" for the insurer may be the paid to date benefit cost of the claim while for the claimant the most important metric is likely his or her lost income. Traditional survival analysis can be helpful here, especially in dealing with open claims, i.e., "right-censored" data (c.f. [2], [4]). Simply taking "t" in the survival analysis models to be paid loss can yield useful reserve estimates (c.f. [4]). workers compensation claims typically involve both medical and wage replacement benefits. Each is expected to follow a distinctive payment pattern that need not be independent of the other. Indeed, that inter-relationship may prove to be a key cost driver. This paper illustrates how a multi-dimensional survival model can reveal those inter-relationships and their cost implications.

Consider, for instance, an issue from the ongoing debate over claim administration protocols. In the workers compensation context, is it better to pursue aggressive medical treatment quickly in an effort to minimize time lost from work, or is it more efficient to spend those resources another way, such as providing job retraining. Clearly the answer may vary tremendously based on the nature of the injury, the age of the worker, the applicable benefit provisions, and a myriad of other considerations.

The main conceptual result of this paper is that traditional survival analysis can be inherently limiting. This is established formally by showing that it is not always possible to define a survival function. The first section of the paper presents a generalization of the survival function to a function of several variables. Many of the basic formulas of survival analysis are readily generalized. The next section discusses censored data and shows how this can introduce new complications in the multi-dimensional context. The concept of a hazard vector field is defined and shown to provide a more general framework than traditional survival analysis. In particular, this framework is capable of dealing with multi-dimensional censored data. It is shown that the existence of a survival function conforms exactly to the "conservative force field" of classical physics. A simple example illustrates how Green's Theorem in the plane applies to comparing and even optimizing paths of action, e.g., as in evaluating claim administration protocols.

The concepts introduced in this paper may lead to the ability to help identify optimum claim handling practices. As noted in the section on further research, much additional work is required to test this approach. Some work that uses this approach to study the resolution pattern of workers compensation back strains shows some promise but is very preliminary. The examples presented here are only numeric illustrations; many have no practical application and some details are left to the reader. Those wishing additional details on the numerical examples or on the application to back strain cases may contact the author.

Section I: Basic Terminology and Notation

Let $\Re^+$ denote the set of nonnegative real numbers and $\Re^n$ denote n-dimensional space. For any $a = (a_1,\dots,a_n) \in \Re^n$, $\Im_a = \{(x_1,\dots,x_n) \mid x_i \ge a_i, 1 \le i \le n\}$; in particular, let $\Im = \Im_0$ denote the "positive quadrant." We regard $\Re^n$ as a model for "space-time" in which each coordinate represents an aging metric. The most natural case is when n=1 and the metric is time. For insurance applications, metrics to keep in mind would be cumulative payments or accumulation of some other quantity associated with claim resolution (e.g. x_1 = time from injury, x_2 = indemnity paid to date, x_3 = medical paid to date, x_4 = ALAE paid to date, etc.). We regard $\Im$ as all possible "failures" or "deaths", all of whose lives begin at the origin. More generally, $\Im_a$ represents the possible future (failure) values subsequent to attaining the point $a \in \Re^n$. Clearly $b \in \Im_a \Leftrightarrow \Im_b \subseteq \Im_a$.

Begin with a continuous probability density function (PDF) of "failures":

$$f : \Im \to \Re^+ \qquad \int_{\Im} f = 1.$$

It is natural to define a survival function as the probability of subsequent failure:

$$S : \Im \to \Re^+ \qquad S(a) = \int_{\Im_a} f = \frac{\int_{\Im_a} f}{\int_{\Im} f} \qquad a \in \Im$$

Observe that f and S uniquely determine one another; indeed, from the fundamental theorem of calculus:

$$f = (-1)^n \frac{\partial^n S}{\partial x_1 .. \partial x_n}.$$

For $b \in \Im_a$, define $f_a(b) = \dfrac{f(b)}{S(a)}$. This defines a PDF on $\Im_a$ in which the origin has been shifted to a and which has survival function $S_a(b) = \dfrac{S(b)}{S(a)}$, the conditional probability assuming survival to a.

Let X be the random variable with PDF f and sample space $\Im$. Because $\Im$ is closed under vector addition (it is an additive semigroup), it is natural to consider the expression:

$$\mu = E(X) = \sum_{x \in \Im} f(x)x$$

as a candidate "expected failure vector". More generally, for $a \in \Im$ this suggests that the expected failure vector for survival beyond a be expressed as:

$$\rho(a) = \sum_{x \in \Im_a} f_a(x)(x-a)$$

This infinite weighted sum, properly interpreted as a limit, can be found (when finite) via integration. Let $\pi_i : \Re^n \to \Re$ denote the usual coordinate projection functions and $\{\varepsilon_i = (0,\ldots 0,1,0,\ldots 0) \mid 1 \le i \le n\}$ the usual set of coordinate unit vectors. Continuity and linearity imply:

$$\rho(a) = \sum_{i=1}^{n} \left(\int_{\Im_a} \pi_i(f_a(x)(x-a)) \right) \varepsilon_i$$

$$= \frac{1}{S(a)} \sum_{i=1}^{n} \left(\int_{a_1}^{\infty} \cdots \int_{a_n}^{\infty} f(x_1, \ldots, x_n)(x_i - a_i)\, dx_1 \ldots dx_n \right) \varepsilon_i$$

The following integration result is a straightforward integration by parts:

Lemma: For any continuous function $g : \Re^+ \to \Re^+$, $b \in \Re^+$ with $\int_0^{\infty} g(t)dt < \infty$

$$\int_b^{\infty} (t-b) g(t) dt = \int_b^{\infty} \int_t^{\infty} g(w) dw dt$$

Proof: Let $b < c \in \Re^+$. Then we have:

$$\int_b^c\int_t^c g(w)dwdt = \int_b^c u\,dv$$

$$u(t) = \int_t^c g(w)dw \qquad v = t - b$$

$$du = -g(t)dt \qquad dv = dt$$

$$= [uv]_b^c - \int_b^c v\,du = \left[(t-b)\int_t^c g(w)dw\right]_b^c + \int_b^c (t-b)g(t)\,dt$$

$$= \int_b^c (t-b)g(t)\,dt$$

and the lemma follows by letting $c \to \infty$.

Define n functions:

$$g_1(t) = \int_{a_2}^{\infty} \dots \int_{a_n}^{\infty} f(t, x_2, \dots, x_n) dx_2 \dots dx_n$$

....

$$g_n(t) = \int_{a_1}^{\infty} \dots \int_{a_{n-1}}^{\infty} f(x_1, \dots, x_{n-1}, t) dx_1 \dots dx_{n-1}$$

Invoking the above lemma and rearranging the order of integration (Fubini's Theorem):

$$\rho(a) = \frac{1}{S(a)} \sum_{i=1}^{n} \left(\int_{a_i}^{\infty} (t - a_i) g_i(t) dt \right) \varepsilon_i$$

$$= \frac{1}{S(a)} \sum_{i=1}^{n} \left(\int_{a_i}^{\infty}\int_{t}^{\infty} g_i(x_i) dx_i dt \right) \varepsilon_i$$

$$= \frac{1}{S(a)} \sum_{i=1}^{n} \left(\int_{0}^{\infty} S(a + t\varepsilon_i) dt \right) \varepsilon_i = \sum_{i=1}^{n} \left(\int_{0}^{\infty} S_a(a + t\varepsilon_i) dt \right) \varepsilon_i,$$

which implies that this candidate for expected survival vector can be determined from conditional survival parallel to the coordinate axes. Note that $\rho : \Im \to \Im$ is a vector field and that:

$$\mu = \rho(0) = \sum_{i=1}^{n}\left(\int_0^{\infty} S(t\varepsilon_i)dt\right)\varepsilon_i$$

Recall that for n=1 the hazard function h can be defined as $h(t) = \dfrac{f(t)}{S(t)}$ or equivalently as $h = -\dfrac{d(\ln(S(t)))}{dt}$. While the first definition readily generalizes to define the *hazard function* $h = \dfrac{f}{S} : \Im \to \Re^+$ for any n, it is the second that is of greater interest. Given a survival function S on $\Im$ the corresponding *hazard vector field* is defined as:

$$\eta = \eta_S : \Im \to \Im \qquad \eta(x) = -\nabla(\ln(S(x))), \quad x \in \Im$$

where ∇ denotes the gradient operator.

For any $a \in \Im$, a *life path of a* is a continuous function $C:[0,1] \to \Im$ satisfying:

$$C(0) = 0$$
$$C(1) = a$$
$$0 \le t \le u \le 1 \Rightarrow C(u) \in \Im_{C(t)}$$

The latter condition simply means that the path progresses into the future. Note that for any $a \in \Im$ and life path C of a, we have:

$$\oint_C \eta = -\ln(S(C(1))) + \ln(S(C(0))) = -\ln(S(a)) \quad \Rightarrow \quad S(a) = e^{-\oint_C \eta}$$

We will, as is often done, occasionally confuse a path C with its image $\{C(t)\}$, implicitly exploiting independence of the line integral to path parameterization.

The traditional language of life contingencies refers to hazard as a "force of mortality". Of course, "force" is inherently a vector concept and the latter expression relates the force of failure η with the probability of survival S. This has a natural appeal as it relates survival to the amount of "work" done traversing a hazardous life path. It gives the term "life work" a new twist and suggests an almost Aristotelian concept of life-giving energy. The existence of a survival function, as defined here, corresponds to the case when the amount of work is independent of the path, analogous to a potential function measuring energy loss in classical physics.

We conclude this section with some simple examples for n=2, in which case we revert to the more conventional xy-plane notation.

Example 1: Let $(a,b) \in \Im$ be a vector in the positive quadrant. The *exponential survival function with parameter vector (a,b)* is defined as:

$$S(x,y)=e^{-ax-by} \qquad f(x,y)=abS(x,y) \quad h(x,y)=ab$$

Note that this models the case of constant expected survival, $\rho(x,y)=(\frac{1}{a},\frac{1}{b})$, and constant hazard field $\eta(x,y)=(a,b)$.

The generalization of Example 1 to $n>2$ is clear. It is not surprising that the expected survival vector is constant exactly when the hazard field is constant. The inverse relationship between the two in that event, $\eta\bullet\rho=n$, has an added geometric appeal since survival is "global" while hazard is "local" (See [3] for a more systematic discussion of the relationship between hazard and expected survival.)

Example 2: After a constant vector field (Example 1), the next simplest vector field is

$\eta(x,y)=c(x,y)$ for some constant $c\in\Re^+$.

For this case,

$$S(x,y)=e^{-c\left(\frac{x^2+y^2}{2}\right)}, \quad f(x,y)=c^2xyS(x,y), \quad h(x,y)=c^2xy$$

We leave to the reader the verification that:

$$\rho(x,y)=\left(\sqrt{\frac{2\pi e^{cx^2}}{c}}(1-\Phi(\sqrt{c}x)),\sqrt{\frac{2\pi e^{cy^2}}{c}}(1-\Phi(\sqrt{c}y))\right),$$

where $\Phi(x)=\sqrt{\frac{1}{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$ is the standard normal cumulative density function.

Example 3: Suppose γ, T, g define another hazard field, survival, and PDF, respectively. Then $\eta+\gamma$ has survival function the product ST and PDF:

$$=S(x,y)g(x,y)+T(x,y)f(x,y)+\int_x^\infty f(t,y)dt\int_y^\infty g(x,t)dt+\int_y^\infty f(x,t)dt\int_x^\infty g(t,y)dt\geq 0.$$

Combining these examples, the "first degree equation" hazard field

$$\eta(x,y)=(a,b)+c(x,y)$$

has the "second degree" survival function $S(x,y)=e^{-\left(ax+by+c\left(\frac{x^2+y^2}{2}\right)\right)}$.

When $n=1$, a hazard function $h(t)$ is often viewed as belonging to a one-parameter family $\{ch(t)\mid c\in\Re^+\}$ of "proportional hazard" functions ([2] considers the mean survival over such a family). A proportional shift $h(t) \mapsto ch(t),\ c \in \Re^+$ in the hazard function corresponds to exponentiation of the survival function $S(t) \mapsto S(t)^c$. The next example shows that this concept becomes more complicated in higher dimensions.

Example 4: The function $S(x, y) = e^{-(x+1)(y+1)}$ is a survival function with PDF $f(x, y) = (xy + x+y)S(x, y)$ and hazard field $\eta(x, y) = (y+1, x+1)$. Letting $T = \sqrt{S}$, we let the reader verify that T is **not** a survival function, as defined here, since it would have PDF:

$$e^{-(x+1)(y+1)}.(\frac{(x+1)(y+1)}{4} - \frac{1}{2}) < 0$$

for (x,y) sufficiently near the origin.

Section II: Censored data and Path Dependence

To make the discussion seem more concrete, let y measure wage replacement benefits and x medical benefits awarded on a workers compensation claim. For convenience, normalize costs so that the interval [0,2] covers the range of feasible amounts. Consider the following table of survival data:

Survival Data			
x	**y**	**Status**	**Count**
1	0	Open	378
1	1	Closed	393
2	2	Closed	229
Total			1,000

In this context "failure" means claim closure, as that corresponds to the end of the life of a claim. Cases open when the information is collected are regarded as censored. The reported values of x and y represent medical and indemnity paid to date figures at that evaluation. For closed cases, the final incurred costs are reported. Consistent with the assumed unit of payment, no case survives beyond (2,2).

Let P_a^b denote the probability of survival from point a to point b. The task is to determine the probability of survival from (0,0) to the point (1,1) = $P_{(0,0)}^{(1,1)}$. Note that there are no observed closures from (0,0) to (1,0) or to (0,1), so we must have $P_{(0,0)}^{(1,0)} = P_{(0,0)}^{(0,1)} = 1$. Since there are 393 failures at (1,1) among 1000 cases, none censored at (0,1), we find that $P_{(0,1)}^{(1,1)} = \frac{1000-393}{1000} = 0.607$. Taking into account the censoring at (1,0), however, implies that $P_{(1,0)}^{(1,1)} = \frac{1000-378-393}{1000-378} = 0.368$. Since $P_{(0,0)}^{(0,1)} P_{(0,1)}^{(1,1)} = 0.607 > 0.368 = P_{(0,0)}^{(1,0)} P_{(1,0)}^{(1,1)}$, this illustrates how censored data leads to a problem determining a probability of survival $S(1,1)$ from (0,0) to (1,1).

The component functions of a hazard vector field η determined from a survival function S are readily expressed in terms of S and the PDF f. For example when n=2 we have:

$$\eta(x,y) = (P(x,y), Q(x,y)) = -\nabla(\ln(S(x,y))$$

$$= -\left(\frac{\partial \ln(S(x,y))}{\partial x}, \frac{\partial \ln(S(x,y))}{\partial y}\right) = \left(\frac{\int_y^\infty f(x,v)dv}{S(x,y)}, \frac{\int_x^\infty f(u,y)du}{S(x,y)}\right)$$

in which the common denominator $S(x,y)$ measures the probability of survival to (x,y) and the numerators the observed "marginal failures" subsequent to (x,y).

In the case of censored data, consider a decomposition:

$$f(x,y) = f_0(x,y) + f_1(x,y)$$

into censored and uncensored observations. Consistent with how (right) censored data is handled in survival analysis when n=1, it is natural to consider

$$\eta_1(x,y) = (P_1(x,y), Q_1(x,y)) = \left(\frac{\int_y^\infty f_1(x,v)dv}{S(x,y)}, \frac{\int_x^\infty f_1(u,y)du}{S(x,y)}\right)$$

in which the numerators measure only observed failures.

Example 5: Begin with:

$$S(x,y) = \frac{1}{(x+1)(y+1)}$$

$$f(x,y) = \left[\frac{1}{(x+1)(y+1)}\right]^2 = S(x,y)^2$$

$$\eta(x,y) = (P(x,y), Q(x,y)) = \frac{1}{2}\left(\frac{1}{(x+1)^2(y+1)}, \frac{1}{(x+1)(y+1)^2}\right)$$

and decompose $f(x,y)$ as:

$$f_0(x,y)=\frac{1+x+y}{(x+1)^3(y+1)^3}$$

$$f_1(x,y)=\frac{xy}{(x+1)^3(y+1)^3}$$

$$\eta_1(x,y)=(P_1(x,y),Q_1(x,y))=\left(\frac{\int_y^\infty f_1(x,v)dv}{S(x,y)},\frac{\int_x^\infty f_1(u,y)du}{S(x,y)}\right)$$

$$=\left(\frac{x(2y+1)}{2(x+1)^2(y+1)},\frac{(2x+1)y}{2(x+1)(y+1)^2}\right)$$

$$\Rightarrow \frac{\partial P_1}{\partial y}-\frac{\partial Q_1}{\partial x}=\left(\frac{x-y}{2}\right)f(x,y)$$

It follows that the vector field η_1 does not have a potential function and in particular does not have the form $-\nabla \ln S_1$ for any survival function S_1. This points out the need to generalize our definitions, as is done in the next section.

Section III: Definition of the Generalized Survival Model

Let $\Gamma=\{C_a \mid a\in\Im; C_a \text{ a life path for a}\}$, $\eta:\Im\to\Im$ a continuous vector field. The corresponding *generalized survival function* $S:\Gamma\to\Re^+$ is determined from

$$S(C_a)=e^{-\oint_{C_a}\eta}$$

The pair η, S is referred to as a *generalized survival model* on $\Im$.

Observe that if η, S and γ, T are generalized survival models, then so is $a\eta+b\gamma, S^aT^b$, for any $a,b\in\Re^+$. In particular, this generalized survival formulation captures situations that cannot be modeled with PDF's, from both this formal arithmetic perspective and as regards the ability to relate survival with choice of life path.

Of course, even for n=1, any continuous function $h:\Re^+\to\Re^+$ can formally define a survival function as $S(t)=e^{-\int_0^t h(w)dw}$ but setting $f(t)=h(t)S(t)$ need not yield a continuous PDF, as considered here. Indeed, $\int_0^\infty f(t)dt=1-p$ where $p=\lim_{t\to 0}S(t)$ can be greater than 0. In that case it is easy to augment *f (t)* by a point mass of probability *p* to achieve a mixed PDF based model. For $n>1$, the relationship between path dependence and the existence of a PDF based model lies somewhat deeper.

Also, for $n=1$, the hazard is interpreted as the instantaneous rate of failure. Consider now the case $n=2$. Following standard convention, express the hazard field as $\eta(x,y)=(P(x,y),Q(x,y))$ and assume also that P and Q are continuously differentiable. Note that for any $t>0$, any life path of $(a+t,b)$ passing through (a,b) has the form $C+D_t$ where C is a life path of (a,b) and $D_t(s)=(a+s,b),\ 0\le s\le t$. It follows that the conditional probability of survival from (a,b) to $(a+t,b)$ is uniquely determined as:

$$e^{-\oint_{D_t}\eta}=\frac{S(C+D)}{S(C)}=p(t)$$

The mean rate of failure $\alpha(t)$ per horizontal unit along D_t is also independent of the choice of C:

$$\alpha(t)=\frac{\left(\dfrac{S(C)-p(t)S(C)}{S(C)}\right)}{t}=\frac{1-p(t)}{t}$$

using the fact that the curve D_t is parameterized by arc length.

We are interested in the instantaneous horizontal rate of failure at (a,b), which is just the limit:

$$\alpha=\lim_{t\to0}\alpha(t)=-\lim_{t\to0}\frac{p(t)-p(0)}{t-0}=-\frac{dp}{dt}\Big|_{t=0}$$

On the other hand:

$$p(t)=e^{-\oint_{D_t}\eta}\quad\Rightarrow$$

$$-\ln p(t)=\oint_{D_t}\eta=\int_{(a,b)}^{(a+t,b)}P(a+x,b)dx+Q(a+x,b)dy$$

$$=\int_0^t P(a+s,b)ds$$

since $dy=0$ along D_t. First differentiating by t and then letting $t\to0$:

$$\frac{1}{p(t)}\left(-\frac{dp}{dt}\right)=P(a+t,b)\quad\Rightarrow\alpha=P(a,b)$$

We conclude that:

$P(a,b)$=instantaneous rate of failure per horizontal age unit at (a,b)

$Q(a,b)$=instantaneous rate of failure per vertical age unit at (a,b).

This is readily generalized to higher dimensions and provides a means to calculate the component functions of the hazard vector field. Clearly the hazard vector field determines a generalized survival function. This discussion shows the converse: a generalized survival function determines conditional probability, whence failure rates parallel to the coordinate axes, which in turn determine a hazard vector field.

For any n and $a \in \Im$, define the curve $D_{a,i,t}(s) = a + s\varepsilon_i, 0 \le s \le t, 1 \le i \le n$. The above discussion on failure rate noted that conditional survival parallel to a coordinate axis is independent of choice of path and the discussion in Section I then suggests the following definition for the expected survival vector

$$\rho(a) = \sum_{i=1}^{n} \left(\int_0^{\infty} e^{-\oint_{D_{a,i,t}} \eta} dt \right) \varepsilon_i \quad for \quad a \in \Im$$

When n=2 and $\eta(x, y) = (P(x, y), Q(x, y))$ the reader can readily verify that

$$\rho(a,b) = \left(\int_0^{\infty} e^{-\int_0^t P(a+s,b)ds} dt, \int_0^{\infty} e^{-\int_0^t Q(a,b+s)ds} dt \right)$$

$$= \left(\int_a^{\infty} e^{-\int_a^t P(s,b)ds} dt, \int_b^{\infty} e^{-\int_b^t Q(a,s)ds} dt \right)$$

Note that for $c>a$:

$$\int_a^{\infty} e^{-\int_a^t P(s,b)ds} dt = \int_a^{c} e^{-\int_a^t P(s,b)ds} dt + \int_c^{\infty} e^{-\int_a^t P(s,b)ds} dt$$

$$\le c - a + e^{-\int_a^c P(s,b)ds} \int_c^{\infty} e^{-\int_c^t P(s,b)ds} dt$$

$$\le c - a + \int_c^{\infty} e^{-\int_c^t P(s,b)ds} dt$$

$$\Rightarrow a + \int_a^{\infty} e^{-\int_a^t P(s,b)ds} dt \le c + \int_c^{\infty} e^{-\int_c^t P(s,b)ds} dt$$

$$\Rightarrow (c,b) + \rho(c,b) \in \Im_{(a,b)+\rho(a,b)}$$

and by symmetry, for $d>b$:

$$(c,d) + \rho(c,d) \in \Im_{(c,b)+\rho(c,b)} \subseteq \Im_{(a,b)+\rho(a,b)}$$

The corresponding result for n=1 is a special case of n=2 and the case n>2 is a straightforward induction using the case n=2. In general, we have:

$$b \in \Im_a \quad \Rightarrow \quad b + \rho(b) \in \Im_{a+\rho(a)}$$

This is intuitively what one would expect and has implications to the task of determining a hazard vector field approximating empirical data (c.f. [3]).

Again, the section concludes with an example:

Example 6: Consider the vector field

$$\eta(x,y) = \left(\frac{y^2}{2}, x^2\right).$$

and consider the following line segment paths:

$$C_1 \text{ from } (0,0) \text{ to } (1,0)$$
$$C_2 \text{ from } (0,0) \text{ to } (0,1)$$
$$C_3 \text{ from } (1,0) \text{ to } (1,1)$$
$$C_4 \text{ from } (0,1) \text{ to } (1,1)$$
$$C_5 \text{ from } (0,0) \text{ to } (1,1)$$

Observe that, with the usual notational conventions, $C_1 + C_3, C_2 + C_4$ and C_5 can be taken as life paths for the point $(1,1)$. For example,

$$\oint_{C_1} \eta = \int_{(0,0)}^{(1,0)} \frac{y^2}{2} dx + x^2 dy \qquad y = 0, dy = 0$$

$$= \int_0^1 0dx + x^2(0) = 0 \quad \Rightarrow S(C_1) = 1$$

The reader can readily verify the following observations:

$$S(C_2) = 1$$

$$S(C_1 + C_3) = \frac{1}{e} \approx 0.368$$

$$S(C_2 + C_4) = S(C_5) = \sqrt{\frac{1}{e}} \approx 0.607,$$

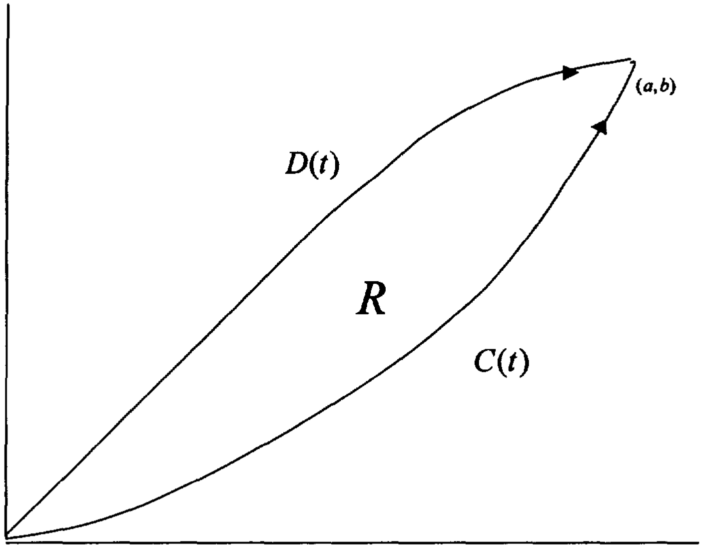
which may explain the rather odd choices for the survival data in the previous section. Note also that

$$\rho(x,y) = \left(\frac{2}{y^2}, \frac{1}{x^2}\right) = \left(\frac{1}{P(x,y)}, \frac{1}{Q(x,y)}\right) \Rightarrow \rho(x,y) \bullet \eta(x,y) = 2.$$

Clearly, a hazard vector field $\eta(x, y) = (P(x, y), Q(x, y))$ and the corresponding expected survival vector field $\eta(x, y)$ should be "inversely related" in some sense. In this example, as in Example 1, their component functions are found to be multiplicative inverses of each other. The interested reader can verify that this is characteristic of the case when $\frac{\partial P}{\partial x} = \frac{\partial Q}{\partial y} = 0$ (c.f. [3]).

Section IV: An Application of Green's Theorem in the Plane

Again consider the case n=2 and let $(a, b) \in \Im$ be a point in the positive quadrant with life paths C and D. We are interested in comparing $S(C)$ with $S(D)$. The case of most interest is when (a,b) is the "first" point beyond the origin at which the life paths meet and so assume further that C lies beneath D in the rectangle $[0,a]\times[0,b]$. The picture is:



We are interested in comparing the probabilities of failure/survival over the two paths. As in the previous section, express the hazard field as $\eta(x,y) = (P(x,y), Q(x,y))$ and assume P and Q are continuously differentiable. Under these conditions, C-D is a closed curve enclosing a simply connected region R. Green's theorem, a topic covered in most advanced calculus courses, relates the line integral over the boundary with an integral over the enclosed region. In this case, it states that:

$$\oint_C \eta - \oint_D \eta = \oint_{C-D} \eta = \iint_R \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\, dx\, dy$$

Letting $r(x,y) = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$—sometimes called the "rotation" of η at (x,y)—it follows that:

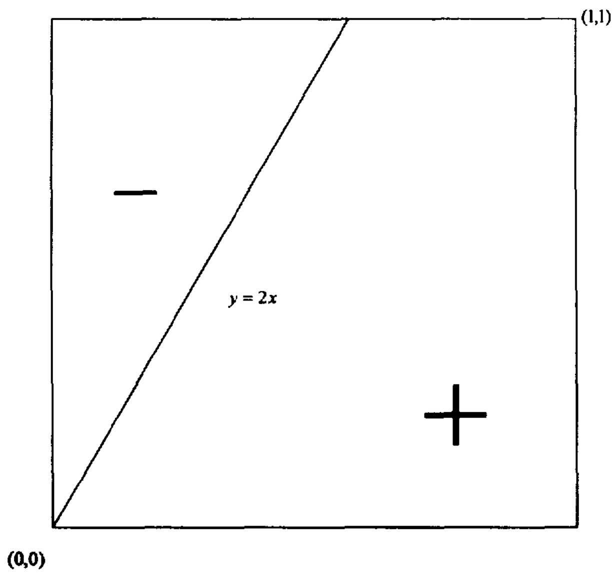
$$S(D) = e^{\alpha} S(C) \quad where \quad \alpha = \iint_R r$$

In particular,

$$r(x,y) \geq 0 \; on \; R \Rightarrow S(D) \geq S(C)$$
$$r(x,y) \leq 0 \; on \; R \Rightarrow S(D) \leq S(C)$$

with strict inequality holding when r does not vanish on R. Clearly, the function $r(x,y)$ is key to the task of identifying paths of least or greatest resistance, i.e., optimum paths for failure or survival.

Example 6 (Continued): Here $r(x,y) = 2x - y$ and as before the focus stays on survival to the point $(a,b) = (1,1)$. All life paths are contained within the unit square where the sign of r is depicted below:



The picture suggests considering the life path defined as:

$$C_6(t) = \begin{cases} (t,2t) & 0 \le t \le \frac{1}{2} \\ (t,1) & \frac{1}{2} \le t \le 1 \end{cases}$$

The reader can readily verify—directly or using Green—that:

$$\oint_{C_6} \eta = \frac{5}{12} \Rightarrow S(C_6) = e^{-\frac{5}{12}} \approx .659$$

Consider a deformation of $C_6 \mapsto \breve{C}_6$ downward that would invade the region for which $r>0$. Taking $\breve{C}_6 = C, C_6 = D$ in the above, we find that $S(C_6) > S(\breve{C}_6)$. On the other hand, any deformation of $C_6 \mapsto \hat{C}_6$ upward would invade the region for which $r<0$. Taking $\hat{C}_6 = D, C_6 = C$ in the above, we find that $S(C_6) > S(\hat{C}_6)$. It follows that the life path C_6 provides the maximum probability of survival to $(1,1)$. A similar argument shows that the life path $C_1 + C_3$ provides the minimum probability of survival to $(1,1)$. Finally, consider, as in Section II, the interpretation when values of x and y represent medical and indemnity benefits paid to date. Subject to this hazard function, the path $C_1 + C_3$ (which corresponds to the "sports medicine" approach of first focusing all resources to medical care) maximizes the probability of claim resolution at $(1,1)$.

It is apparent from the example that optimal paths can be expected to trace along solutions to $r(x,y)=0$ and the boundary of the rectangle. Observe that in the interpretation of Example 6, time was not included among the coordinates. Instead, time was relegated to the role of parameter of life paths. That is appropriate provided the focus is more on costs than on their specific timing. If, for example, it is desired to estimate expected time to failure, it would make sense to include time among the coordinates and look particularly at the expected survival vector component in that direction. Similarly, if the timing of payments is at issue, such as with claim administration protocols, it is natural to explicitly include time as a coordinate in the model. Given the way data is collected, time stamped payment information is the most natural source for capturing a life path and time is the most natural parameter.

Green's theorem comes neatly into play when considering alternative paths for getting to the same place, i.e., when resources are already allocated and it is a question of optimizing their effect on claim resolution. Logically prior to this, of course, is the issue of allocating resources, as illustrated in yet another revisit to the example:

Example 6 (Continued): Suppose we have fixed resources $\beta > 0$ and we consider the portion $\alpha \in [0,1]$ to be spent on medical. Clearly this involves considering life paths to the line $x + y = \beta$. So let $C_{\alpha,\beta}$ denote the linear life path from (0,0) to $(\alpha\beta, (1-\alpha)\beta)$. We leave to the reader the verification that:

$$\gamma(\alpha,\beta) = \oint_{C_{\alpha,\beta}} \eta = \frac{(\alpha - \alpha^3)\beta^3}{6}$$

It follows that for any $\beta > 0$, $\gamma(\alpha,\beta)$ has a relative maximum at $\alpha = \sqrt{\frac{1}{3}}$ and so allocating that portion of every dollar to medical would follow along the straight path

$$\left\{ \left(\sqrt{\frac{1}{3}}t, \left(1 - \sqrt{\frac{1}{3}}\right)t \right) \mid t \geq 0 \right\}$$

that maximizes the probability of resolving the claim.

There is also the converse issue, suppose you are confronted with a claim that requires a certain amount of work to close, how can you minimize the cost outlay? This related allocation problem is illustrated in our final revisit to the example:

Example 6 (Concluded): Suppose we have a fixed amount of work $\beta > 0$ needed to close a claim and we wish to find a life path that requires the least possible total payment $x + y$. We simplify the problem and only consider straight-line solutions and let α denote the slope. Let C_a denote the linear life path from (0,0) to $(a, \alpha a)$. The reader can verify that our constraint translates into:

$$\beta = \oint_{C_a} \eta = \int_0^a \frac{\alpha^2 u^2}{2} + \alpha u^2 du = \left(\frac{\alpha^2}{2} + \alpha\right)\left(\frac{a^3}{3}\right)$$

and that the outlay $a + \alpha a$ is minimized when $\alpha = \sqrt{3} - 1$. We find that in the most cost-effective solution, the (constant) portion spent on medical $= \frac{\sqrt{3}-1}{\sqrt{3}}$ is independent of the fixed amount of work β required to close the claim.

We conclude this section with a formulation of Green's theorem suitable for comparing survival along any two life paths C and D of $(a,b) \in \Im$. For any $x \in [0,a]$ let $L_x = \{(x,t) \mid t \in [0,b]\}$ be the vertical line segment above x. Our assumptions imply that:

$$L_x \cap D = \{(x,t) \mid t \in [d_1(x), d_2(x)]\}$$
$$L_x \cap C = \{(x,t) \mid t \in [c_1(x), c_2(x)]\}$$

And we may define:

$$\delta(x,y) = \begin{cases} -1 & d_2(x) < c_1(x) \\ +1 & c_2(x) < d_1(x) \\ 0 & otherwise \end{cases}$$

Pictorially, δ is 1 when C lies below D and -1 when D lies below C, in effect flagging the two possible orientations the life paths can traverse around the region R they enclose. All life paths to (a,b) lie in the closed rectangle $[0,a]\times[0,b]$ and the path:

$$\hat{C}(t) = \begin{cases} (0,2bt) & 0 \le t \le \frac{1}{2} \\ (2a(t-\frac{1}{2}),b) & \frac{1}{2} \le t \le 1 \end{cases}$$

is the "top" top life path. Let

$$R_1 = \text{simply connected region enclosed by } C\text{-}\hat{C}$$

$$R_2 = \text{simply connected region enclosed by } D\text{-}\hat{C}$$

$$R = (R_1 \cup R_2) - (R_1 \cap R_2)$$

By Green's theorem:

$$\oint_{C-D}\eta = \oint_{C-\hat{C}}\eta - \oint_{D-\hat{C}}\eta = \iint_{R_1} r - \iint_{R_2} r = \iint_{R} \delta r$$

$$\Rightarrow S(D) = e^{\alpha}S(C) \quad where \quad \alpha = \iint_{R} \delta r$$

This provides a general comparison formula that is amenable to numeric evaluation. In practice, though, a simple chart of the sign of $r(x,y)$ over the applicable rectangle is the best starting point. The key, therefore, to identifying optimal paths is a representation of η that yields a sufficiently accurate picture of r.

Section V: Further Research

The question remains how to determine a hazard field from empirical data. One simple approach is to restrict the domain of the function to regions over which the hazard vector is assumed constant and then approximate it by estimating the coordinate failure rates. For this, traditional survival analysis methods suffice. SAS, for example, is well suited since its survival analysis procedures can be performed over cells of data and its time variable can be set to measure changes parallel to the coordinate variables (see [1]). General curve fitting techniques can then be used to paste the pieces together. Clearly a more systematic approach, especially a computer algorithm, would be useful. An alternative is to first estimate the expected survival vector field ρ—which is more straightforward in concept—and then "invert" that field in some fashion to derive the hazard vector field η (this is considered in [3]).

A generalized survival model can be used to assign a case reserve "vector". Unlike traditional reserve formulas, the vector would account for the interaction of component cost liabilities. Properly formulated, it would provide integrated benchmarks for both the prospective duration and various dollar costs of a claim. Note that the definition of expected survival vector field presented here is strictly prospective. It would be interesting to see whether the theory can yield a "tangent reserve vector" (or higher derivative vectors) defined on life paths and sensitive to the prior history of the claim.

It would also be interesting and potentially very valuable to determine whether an insurer has any tendency to follow paths of "greatest or least resistance" in resolving cases. The ability to identify optimum paths might eventually yield valuable information on protocols for case management. Example 6 is indicative of how to exploit Green's Theorem in such an investigation, not to mention first semester calculus.

Example 4 illustrates that the concept of a proportional hazard relationship becomes more complicated in higher dimensions. Indeed, the concept itself can be blown up n^2-fold from scalar to matrix multiplication. Further research is needed to determine what concepts work best. The Cox proportional hazard model (see [1]) is the standard tool for relating explanatory variables ("covariates") with the hazard function. Because each component along a life path implies essentially the same failure **sequence**, the Cox model will typically associate the same covariate proportional shift irrespective of which coordinate x_i is chosen as the time t variable. Alternatively, a parameter for the life path could be used as time t. As a result, the Cox model can be used in this context but only with the understanding that the proportional effect is assumed to be uniform over all values of all components. By the same token, so-called "time dependent" interventions can also be analyzed using the Cox model provided the intervention is consistently defined among the n components. This should not be a problem with time-stamped data where time is used to parameterize the life paths.

Of particular value would be a generalization of the Cox model approach that avoids such strong "inter-dimensional" assumptions on constant proportionality. The ideal solution would be the ability to model covariate impact on the hazard vector field via pre or post multiplication by a constant matrix. Presumably, determining the "best" such matrix would involve constructing appropriate maximum likelihood functions. The discussion in Section III, however, suggests that this may not be straightforward.

Sometimes all of $\mathfrak{I}$ may exceed the "natural" sample space for a particular problem. A subset (e.g. manifold as in [5]) might be more appropriate and the "Stokes type" theorems may prove useful in that context, analogous to the use of Green's theorem in the simple example discussed here. Applications of "advanced calculus" have traditionally been the purview of physicists and engineers, not actuaries. Use of multivariate survival models may help level that playing field.

References

[1] Allison, Paul D., *Survival Analysis Using the SAS® System*: A Practical Guide, The SAS Institute, Inc., 1995.

[2] Corro, Dan, *Calculating the Change in Mean Duration of a Shift in the Hazard Rate Function*, to appear in CAS Forum, Winter 2001.

[3] Corro, Dan, *A Note on the Inverse Relationship of Hazard with Life Expectancy*, in preparation.

[4] Corro, Dan, *Modeling Loss Development with Micro Data*, CAS Forum, Fall 2000.

[5] Spivak, Michael, *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*, Perseus Books, 1965.

Surplus Allocation: A DFA Application

Kevin J. Olsen, FCAS, ASA, MAAA, CFA

BIOGRAPHY

Kevin J. Olsen is a Fellow of the Casualty Actuarial Society, Associate of the Society of Actuaries, Member of the American Academy of Actuaries, and a Chartered Financial Analyst charterholder. Mr. Olsen is employed by GuideOne Insurance in West Des Moines, Iowa where he is active in reserving issues and financial projects. He received his B.S.B.A in Actuarial Science from Drake University. Mr. Olsen is an active Casualty Actuarial Society member currently serving on the Examination Committee.

ACKNOWLEDGEMENTS

The author would like to thank David A. Withers for his encouragement and advice. He also contributed much appreciated time and assistance reviewing, editing, and suggesting improvements to this paper. The author would also like to recognize Thomas E. Hettinger for a final read through and critique.

Surplus Allocation:

A DFA Application

Kevin J. Olsen

ABSTRACT

Surplus allocation has been requested from actuaries many times over the years. There are those who feel surplus allocation of any sort is incomprehensible. Since actuaries are asked to allocate surplus, we need to ensure the processes being used are sound. It is such a request from upper management that sent the author looking for the methods employed by others and pondering what additional methods could be constructed. This paper reviews reserve and duration based allocation methods and then ventures into devising an alternative method based on variation. A brief discussion is also included on what surplus amount should be used.

PURPOSE

The purpose of this paper is to share methods for surplus allocation with others, receive feedback on these methods, and promote further development. The author is a company actuary in the pursuit of answers for management. This project was begun to answer a question presented by the company's CFO. The questions raised were non-actuarially based but needed to be answered by someone with a financial understanding. Given the company's surplus, what is the optimal distribution of surplus by line of business? This will allow tracking, calculating, and determining profitability of each line of business on its own.

INTRODUCTION

This paper will review and analyze three methods of allocating surplus. The methods can be used to distribute current surplus by line of business. This is desired for many reasons including pricing activities, determining ROE by line of business, figuring premium-to surplus ratios by line of business, and distributing investment income to line of business.

Although many ways have been discussed to allocate surplus, there is no single standard accepted by everybody. California Proposition 103 used the proportion of loss and unearned premium reserve to allocate surplus. It has been suggested that surplus being used for pricing purposes should be allocated based on one's favorite risk load formula[1]. Other methods include allocating surplus in proportion to loss reserves, in proportion to duration, or based on the coefficient of variation in loss ratios. This paper will start with the simpler methods and venture into a variance-based method. The methods will discuss allocation by reserves, duration, and variation.

Keep in mind the allocation of surplus to line of business will not mean line of business independence, because the total amount of the surplus is still there to support the company as a whole. The standard deviation of the enterprise surplus or operating gain will always be less than the sum of the standard deviations by line of business, due to less than perfect correlation between the lines of business.

[1] Suggested by Glenn G. Meyers via the CASNET.

METHODS

REVIEW OF SURPLUS

What is the purpose of surplus? Surplus is there for two purposes 1) to support insurance company operations and 2) to support other activities. The surplus allocation for the purposes of this paper deals with supporting insurance company operations. This is an amount necessary to cover risks such as the variation in liabilities at a point in time, as well as prepare for future needs. From a statutory view, as the company grows the expenses are realized immediately, while the premiums are earned over the course of the policy. If the company accelerates its growth, there will be a reduction of surplus to cover the current expenses. From a going-concern basis, the future liabilities also need to be considered in the surplus allocation.

> "Surplus [exists to] protect the insurer against several types of risk. Asset risk is the risk that financial assets will depreciate (e.g., bonds will default or stock prices will drop). Pricing risk is the risk that at policy expiration, incurred losses and expenses will be greater than expected. Reserving risk is the risk that loss reserves will not cover ultimate loss payments. Asset-liability mismatch risk is the risk that changes in interest rates will affect the market value of certain assets, such as bonds, differently than that of liabilities. Catastrophe risk is the risk that unforeseen losses, such as hurricanes or earthquakes, will depress the return realized by the insurer. Reinsurance risk is the risk that reinsurance recoverables will not be collected. Credit risk is the risk that agents will not remit premium balances or that insureds will not remit accrued retrospective premiums." [5]

RESERVE METHOD

Distributing surplus based on loss reserves and unearned premium reserves may be the easiest method. Allocating surplus according to the volume of business per line is a logical choice since surplus is committed when the policy is written and released when the loss is paid. If it is a stable book of business, the loss reserves and unearned premium reserves will remain relatively constant from year to year. California's proposition 103 used this method to allocate surplus to line of business for their calculations.

This method matches available surplus to line of business in proportion to reserves held. There are no tricky calculations or multiple iterations. The necessary information can be found in the annual actuarial report or the annual statement.

The method begins by listing the ultimate loss reserves needed by line of business and summing them for the enterprise. The same is done for the unearned premium reserve. These are shown in columns 1 and 2 of Table 1 below, while the sum is shown in column 3. For each line of business, take the respective reserve sum and divide by the enterprise sum. This gives the distribution of reserves by line of business, which can be applied to surplus. (See Table 1 or Exhibit 1.) Enterprise surplus can then be multiplied by the corresponding percentages to get the amount of surplus by line of business.

Table 1

	(1) Loss Reserves (000's)	(2) Unearned Premium (000's)	(3) Sum (000's)	(4) Dist.	(5) Surplus (000's)
Homeowners	66,900	27,277	94,177	7.3%	36,500
Personal Auto Liability	385,914	44,801	430,715	33.5%	167,500
Personal Auto Phys Dam	37,426	41,044	78,470	6.1%	30,500
Commercial Auto Liability	112,318	6,093	118,411	9.2%	46,000
Commercial Auto Phys Dam	3,599	2,218	5,817	0.5%	2,500
CPP Liability	141,808	51,320	193,128	15.0%	75,000
CPP Property Damage	63,106	68,577	131,683	10.2%	51,000
Other Liability	3,725	7,565	11,290	0.9%	4,500
Umbrella	1,394	316	1,710	0.1%	500
Workers Compensation	146,415	74,058	220,473	17.2%	86,000
Enterprise	962,607	323,269	1,285,876	100.0%	500,000

The reserve method is a quick and easy method to use, but there are several disadvantages to using this method. It does not consider the length of the reserve pay-out tail, adjustments in the reserve payment pattern, or the time value of money. All of these can cause variations or unexpected results, the precise thing surplus is there to cover. This is a static method. The distribution is determined based on an expected value at a point in time and does not consider future changes in the distribution by line of business.

The reserve method of allocation considers only the pricing and reserving risks. Larger amounts of surplus are allocated to the lines of business holding larger reserves. This method ignores the five other significant areas of variability referenced above in determining the surplus allocation. These five neglected risks include asset risk, asset-liability mismatch risk, catastrophe risk, reinsurance risk, and credit risk.

For more information on different reserve based methods reference "An Evaluation of Surplus Methods Underlying Risk Based Capital Calculations" by Michael Miller and Jerry Rapp (1992 Discussion Paper Program, Vol. 1).

DURATION METHOD

Many people on the CAS web site and CASNET touted duration as a means to allocate surplus. Duration allocation is perceived to be superior to loss and unearned premium reserve allocation since duration considers payment pattern changes and interest rate changes in the duration calculation. Longer tail lines receive relatively more surplus to cover the larger potential volatility in the payment pattern.

Duration is a time value weighted pay-out length. In other words, duration is a weighted average term to completion where the years are weighted by the present value of the related cash flows. [6]

$$\text{Duration} = \frac{\sum_{t=1}^{n} [\,(t \cdot CF_t)/(1+y)^t\,]}{\sum_{t=1}^{n} [\,CF_t/(1+y)^t\,]}$$

CF_t = Cash Flow in year t
y = yield to maturity
t = year of cash flow
n = number of years to maturity

Duration Example

Table 2

Calendar Year	(1) Projected Paid	(2) Year	(3) Present Value @ 6.5%	(4) (2) x (3) Weighted PV @ 6.5%
1997	350,000	0.5	339,151	169,576
1998	210,000	1.5	191,071	286,607
1999	60,000	2.5	51,260	128,150
2000	20,000	3.5	16,044	56,153
2001	7,500	4.5	5,649	25,421
Total	647,500		603,175	665,907

Macaulay Duration = 665,907 / 603,175 = 1.104

Modified Duration = 1.104 / 1.065 = 1.0366

Table 2 gives an example of a duration calculation. Column 1 is the amount projected to be paid in each calendar year. This includes payments from all accident years 1997 and prior. Column 2 shows that the duration is being examined from the beginning of 1997 since the average payment is expected to be paid half way through the year, assuming that in any calendar year the payments are uniform. The present value of column 1 at 6.5% is shown in column 3, while column 4 is (2) times (3). Column 4 gives a weighted present value based on the length until payment. The Macaulay duration is the sum of column 4 divided by the sum of column 3. The Modified duration is the Macaulay duration divided by 1 plus the interest rate used.

The duration method can be applied using a Dynamic Financial Analysis (DFA) model that incorporates changing discount rates, payment patterns, and inflation amounts by iteration in the calculation. Dynamo2 is an Excel based model developed by the actuarial consulting firm of Miller, Rapp, Herbers & Terry (MRH&T) used by the author. Further description of the model can be found in Appendix A.

The DFA model needed some programming additions to capture and calculate the necessary components for duration. Appendix B lays out the changes made to generate the payments by accident year and calendar year and to generate the interest rate.

With the necessary information obtained, formulas were inserted in the DFA model to calculate duration as shown in the example above. A sample iteration of this duration process for the homeowners line of business is shown in Exhibit 2. After the DFA ran 1,000 iterations (maximizing computing capacity), durations were selected equal to the means of the 1,000 durations by line of business.

Table 3 below shows the process of the duration method. Determining a distribution of surplus begins by normalizing the duration by line of business with the enterprise duration. Each line of business duration in column 1 is divided by the enterprise duration. Multiplying the resulting relativities in column 2 by the inverse of the company's premium-to-surplus ratio changes the relativities to the amount of surplus needed per dollar of premium. The next step is to apply the appropriate premium from column 4 to each line of business to arrive at the estimated surplus in column 5. From here a distribution may be determined by dividing line of business estimated surplus by the enterprise surplus. The resulting distribution in column 6 can then be used to spread the real surplus to line of business. This method is also outlined in Exhibit 3.

Table 3

	(1) Avg. Duration	(2) Relativity	(3) Surplus-to-Prem	(4) Premium (000s)	(5) Est. Surplus (000s)	(6) Dist.	(7) Surplus (000's)
Homeowners	1.8677	0.8309	0.4985	54,553	27,197	5.8%	29,000
Personal Auto Liab	2.0180	0.8978	0.5387	179,204	96,531	20.5%	102,500
Personal Auto Phys Dam	0.9251	0.4115	0.2469	164,175	40,539	8.6%	43,000
Comm'l Auto Liab	2.5823	1.1488	0.6893	24,370	16,798	3.6%	18,000
Comm'l Auto Phys Dam	1.3161	0.5855	0.3513	8,872	3,117	0.7%	3,500
CPP Liab	3.2253	1.4349	0.8609	102,640	88,367	18.7%	93,500
CPP Property	1.7051	0.7586	0.4551	137,154	62,424	13.2%	66,000
Other Liability	2.2799	1.0143	0.6086	15,130	9,208	1.9%	9,500
Umbrella	2.2278	0.9911	0.5947	631	375	0.1%	500
Workers Comp	3.2076	1.4270	0.8562	148,116	126,817	26.9%	134,500
Enterprise	2.2478	1.0000		834,844	471,373	100%	500,000

Prem / Surplus = 1.67 (Assumed)[2]
Surplus / Prem = 1 / (Prem / Surplus) = 0.60

(2) = (1) / [(1) Enterprise]
(3) = (2) * 0.60
(5) = (3) * (4)
(7) = 500,000 * (6)

In this presentation the Macaulay duration was used. The question may come up as to why use the Macaulay duration and not the Modified duration. Since this method deals with relative duration by dividing each line of business by the enterprise duration, it does

[2] The correct premium to surplus ratio is assumed to be 1.67.

not matter which one is used. If the modified duration were used, then all the durations would be divided by the same factor maintaining the same relativities between them.

In addition to the advantages listed above, there are a few disadvantages to using the duration method to allocate surplus. The duration method distributes surplus based on projected ultimate losses for past years and the payment pattern for those years. From that point of view, using duration has a run-off view point. It allocates surplus to lines of business in relation to how those lines will run-off and in relation to current premium volume. This covers the vulnerability to greater variation in the longer payout lines of business. This is a static view of business at a point in time. It does not consider future growth or changes in the mix of business going forward. For a company that plans on continuing to write business and grow, this might not be the best option. Surplus needs to be allocated for future premium growth. For statutory accounting, the expenses of writing policies are recognized immediately, while the premiums are earned over the course of the policies. This is why companies with accelerated growth may see surplus decline (statutory surplus, not market value surplus). "Rapid premium growth precedes nearly all of the major failures. Rapid growth is not harmful, per se. However, rapid premium growth reduces the margin for error in the operation of insurers." [4] Additional surplus is needed to cover the reduced margin of error for growing lines of business.

The duration method does a better job than the reserve method of considering the risks surplus is to protect against. The lines of business with longer pay-out patterns have higher durations. Here duration is a proxy for the riskiness of the long tail lines. Longer tailed lines are exposed to more interest rate and payment pattern changes incorporated in reserving risk, asset risk, and asset-liability mismatch risk. By using duration as a proxy, this allocation method covers these risks. Again, four risks surplus is to protect against are not even considered by this method (pricing risk, catastrophe risk, reinsurance risk, and credit risk).

Keep in mind that even though this model considers variation in the payment pattern, judicial or legislative changes that could effect payments are not considered. Such changes would create greater variability, but are difficult if not impossible to predict. These types of changes can not be foreseen on any method presented here.

VARIATION METHOD

The variation method is one that the author developed while working with the DFA model and trying to answer the CFO's questions. It is a forward-looking method on what may happen. Loss reserves are already set up to cover losses that have occurred. Surplus exists for unexpected events or variations from the norm. This method uses standard deviations on a comparable basis among lines of business to distribute surplus.

The variation method uses the calendar year operating gain by line of business from each iteration of the DFA, calculated by adding net underwriting profit to the investment return during the calendar year. To calculate such information, additions needed to be made to the DFA spreadsheet to capture interest earned by line of business. This is described in Appendix C.

Following the steps described in Appendix C, the investment return by line of business to be included in the operating profit was derived as the amount of reserves available for investment times the rate of return for the appropriate year. The calendar year net operating profit was calculated by adding this investment return and the calendar year net underwriting profit by line of business.

The next step was to compare the variation between lines of business. Using the variance of operating profit alone would give results that are difficult to compare between lines of business. Each line of business variance would be based on differing amounts of premium and number of policies. To put all lines of business on a comparable basis the operating gain needed to be normalized before determining the variance.

The net operating gain was divided by the net written premium for that line of business. This ratio is a unit of measure with the dollar units canceling out. This put all lines of business on a net operating gain per dollar of net written premium basis before the variance was determined.

As the steps of the variation method calculation are described, reference will be made to the portions of Exhibit 4 discussed in the text. Exhibit 4 shows this method laid out in its entirety. By capturing the operating gain by line of business for each calendar year, the

@Risk software calculates the standard deviation of each line and year over the 1,000 iterations.

Table 4 shows the results of the simulation. Columns (1) through (5) are the standard deviations of net operating gain per dollar of net written premium. This information was generated by the DFA model and @Risk. Appendix D lays out the credibility weightings of these standard deviations.

Table 4

Standard Deviation of Net Operating Gain
Per Dollar of Net Written Premium

	(1)	(2)	(3)	(4)	(5)
	1998	1999	2000	2001	2002
Home	0.4532	2.2769	0.6970	0.6594	0.4550
P Auto Liab	0.1574	0.1235	0.1461	0.1395	0.1536
P Auto Phys Dam	0.1199	0.9526	0.2524	0.2326	0.1184
C Auto Liab	1.3752	1.4673	1.6137	1.6697	1.7350
C Auto Phys Dam	0.4705	2.3967	0.7265	0.6892	0.5526
CPP Liab	0.0575	0.0646	0.1012	0.0772	0.0841
CPP Prop	0.5931	4.5416	1.1913	1.0898	0.5455
Other Liab	0.1688	0.2055	0.2183	0.2339	0.2391
Other Liab - Umbrella	0.8574	0.9288	1.1948	1.2197	1.3292
Workers Comp	0.1308	0.1456	0.1748	0.1856	0.2019
Personal	0.1255	0.7125	0.2001	0.1830	0.1159
Commercial	0.2088	1.4789	0.3995	0.3605	0.2009
Enterprise	0.1544	1.1122	0.2955	0.2666	0.1367

Table 5 below shows the remaining steps to determine the surplus allocation of the variation method. Dividing the credibility weighted standard deviations (Table 8, column 12, Appendix D) by the average standard deviation of the enterprise (Table 8, column 13, Appendix D) normalizes the credibility-weighted standard deviations. Multiplying the resulting relativities by the inverse of the company's premium-to-surplus ratio changes the relativities to the amount of surplus needed per dollar of premium. The next step is to apply the appropriate premium to each line of business to arrive at the estimated surplus (column 17 = (15) * (16)). Appropriate premium could include the year-end premium by line of business or the first year's projected premium. From here a distribution may be determined by dividing line of business estimated surplus by the enterprise estimated surplus. The resulting distribution in column 18 can then be used to spread the real surplus to line of business.

Table 5

	(12) Credibility Weighted Std Dev	(14) Relativity	(15) Surplus-to-Prem	(16) Premium (000s)	(17) Est. Surplus (000s)	(18) Dist.	(19) Surplus (000s)
Homeowners	0.8621	2.1932	1.3133	54,553	71,645	13.0%	65,000
Personal Auto Liab	0.1441	0.3665	0.2195	179,204	39,331	7.1%	35,500
Personal Auto Phys Dam	0.3569	0.9079	0.5436	164,175	89,252	16.2%	81,000
Comm'l Auto Liab	1.5628	3.9757	2.3807	24,370	58,017	10.5%	52,500
Comm'l Auto Phys Dam	0.9035	2.2985	1.3763	8,872	12,221	2.2%	11,000
CPP Liab	0.0770	0.1959	0.1173	102,640	12,041	2.2%	11,000
CPP Property	1.0756	2.7363	1.6385	137,154	224,728	40.8%	204,000
Other Liability	0.2133	0.5427	0.3250	15,130	4,917	0.9%	4,500
Umbrella	1.0980	2.7933	1.6727	631	1,055	0.2%	1,000
Workers Comp	0.1680	0.4273	0.2559	148,116	37,899	6.9%	34,500
Enterprise	0.3931	1.0000		834,844	551,095	100%	500,000

Prem / Surplus = 1.67 (assumed)[3]
Surplus / Prem = 1 / (Prem / Surplus) = 0.60

(15) = (14) * 0.60
(17) = (15) * (16)

[3] The correct premium to surplus ratio is assumed to be 1.67.

The distribution created by the variation method may raise some questions. Why is it that the property and physical damage coverages receive more surplus based on this method? The property and physical damage coverages are subject to catastrophes and therefore more variation from year to year. The variation is a result of both the frequency and the severity of catastrophes. The liability lines have the potential for high single occurrence pay-outs by policy, but the number of these are relatively consistent from year to year. The law of large numbers makes predicting the result for this line of business more consistent.

An example to look at is the amount of surplus allocated to Commercial Auto Physical Damage (CAPD) and CPP Liability (CPP Liab). As can be seen in Table 5, both of these lines are allocated $11,000,000 surplus, whereas the premium for CPP Liab is 11.5 times as large as that for CAPD. In CPP Liab, the law of large numbers helps smooth results, while CAPD is subject to catastrophes. The reinsurance in place also underlies these results.

Both liability and property lines of business for smaller companies are affected by variations in large losses from year to year. The author did not test which lines of business had more variability in large losses, attributing the major variations between lines of business to catastrophes. Changing reinsurance agreements or types of business written could reduce the impact of catastrophe losses.

This method contains many of the characteristics that are desired from a surplus allocation method. The length and amount of the tail are considered with varying payment patterns, incorporating reserving risk. The DFA model varies interest rates to include asset risk into the considerations. The varying interest rate is brought into the operating gain through the investment income. Using operating gains also reflects pricing risk by including variability of loss ratios embedded in the operating return and catastrophe risk by including simulated catastrophes in the underwriting results. The impact of asset-liability mismatch risk is included by varying the interest rates in the model as well as varying the ultimate loss and payment patterns included in the operating gains. The DFA model does not consider reinsurance risk or credit risk, but these could be incorporated based on distributions of uncollectability. Reinsurance risk may be considered negligible depending on the reinsurers' A.M. Best ratings.

This method goes beyond the first two methods and looks to the future. This is a going-concern method, which tries to reveal what the distribution by line of business should be going forward. To do this it incorporates the company's growth plans by line of business and the variability by line of business based on the growth plans and past experience. If the company is going to cut rates to grow more, then this is included in the variation in net operating gain per dollar of net written premium and figured into the standard deviation. Most company changes in growth, mix of business, or type of business are reflected in the operating gains as long as the DFA model is set up appropriately to reflect these changes.

Non-catastrophe reinsurance levels also influence the variability by line of business. On a net of reinsurance basis, as the threshold for excess of loss coverage is reduced, the variability of results also declines. "[A]ny risk which lowers the aggregate Exposure Ratio of the portfolio has added capacity to the portfolio."[9] The exposure ratio is the coefficient of variation (standard deviation / mean). As the variability decreases, the level of surplus needed for the line of business decreases freeing up surplus for other uses.

COMPARISION OF METHODS

The three different methods presented give widely varying results as can be seen in Table 6 below.

Table 6

	(1) Reserve Method	(2) Duration Method	(3) Variation Method	(4) Driving Risks
Homeowners	7.3%	5.8%	13.0%	Catastrophe risk
Personal Auto Liability	33.5%	20.5%	7.1%	Pricing & Interest Rate
Personal Auto Phys Dam	6.1%	8.6%	16.2%	Catastrophe risk
Commercial Auto Liability	9.2%	3.6%	10.5%	Pricing & Interest Risk
Commercial Auto Phys Dam	0.5%	0.7%	2.2%	Catastrophe risk
CPP Liability	15.0%	18.7%	2.2%	Pricing & Interest Risk
CPP Property Damage	10.2%	13.2%	40.8%	Catastrophe risk
Other Liability	0.9%	1.9%	0.9%	Pricing & Interest Risk
Umbrella	0.1%	0.1%	0.2%	Pricing & Reinsurance
Workers Compensation	17.2%	26.9%	6.9%	Pricing & Interest Rate
Enterprise	100.0%	100.0%	100.0%	Catastrophe & Pricing risk

These variations are the result of the reasoning behind the methods. Looking at personal auto liability, the loss reserve method in column (1) allocates 33.5% to this line, whereas, the duration method apportions 20.5% and the variation method only 7.1%. The personal auto liability line of business has a consistent amount of losses every year and consistent sales growth producing higher reserves held. The reserve method reflects this explicitly. The duration method analysis notes that the payout pattern is weighted heavily to the earlier years. This does not allow much time for adverse development. The variation method looks at the reserves and the payout pattern, but also considers that from year to year the loss ratios are consistent. The ultimate personal auto liability losses can be reasonably estimated from year to year without much variation from expected. Therefore less surplus would be needed for unforeseen circumstances.

The homeowners line of business is another good example. With the payouts being quick and settlements rather fast, the level of reserves carried is relatively low. The reserve method looks only at the carried reserves to determine the allocation (7.3%). The duration method considers that the pay-out pattern is relatively short meaning less surplus is necessary (5.8%). Yet when losses are compared from year to year there is greater variation due to catastrophes. The surplus necessary to cover these greater variations is 13.0%.

The driving risks affecting surplus differs for each line of business. For example, catastrophes have more of an impact on property lines than liability coverages. There are certain sets of risks for each line of business that maintain significant influence on results.

These driving risks are listed in column 4 of Table 6.

SURPLUS

What overall amount of surplus should be used? All of the methods discussed above allocate a stated surplus amount. There are a few different methods to determine how much surplus is to be distributed.

ACTUAL

The most straightforward method would be to use the company's actual surplus as of year-end. This amount would then be distributed back to line of business based on the method of choice. A few problems with this method would be if the company was over capitalized (under capitalized). If this were the case than too much (little) would be allocated. As stated toward the beginning of the paper, surplus is there to support insurance operations as well as other activities. The surplus to allocate should be the amount supporting the insurance operations.

Actual surplus also has many definitions to consider. If allocating actual surplus, is it market value, statutory value, or GAAP value? Should equity in the unearned premium reserve or the discounted amount from the loss reserve discount factor be included?

PREMIUM-SURPLUS RATIO

A second method pegs the surplus at a certain premium to surplus ratio (P/S). There are a variety of reasons and justifications for selecting a certain P/S ratio. The P/S ratio could be selected by management's desire not to exceed a certain P/S ratio, say 2:1. It could be pegged to match a certain peer group in the industry. A word of caution: P/S ratios can be manipulated from company to company.

OPERATING GAIN DISTRIBUTION

The amount of surplus needed by a company is based on its aversion to risk. Assume that a company's risk manager determines that they want to be 95% confident that the surplus doesn't decrease, or 90% confident the surplus decreases by no more than 10%. To do this the company would need to generate a distribution of the change in surplus for the year. Another alternative would be to use operating gain for the year. In both of these distributions the desired confidence level amount is found by referencing (1.00 – confidence level) corresponding to the cumulative percentages for the distribution. By choosing the corresponding amount from the distribution, it can be used to determine the desired amount of surplus for the company.

The net operating gains from the DFA model iterations used in the allocation process can be captured and set into a distribution. Using the example from above, the goal would be 90% confidence that surplus would decrease at most 10% in the given year. From the

1,000 iterations, the Enterprise operating gain for 1998 was captured with the resulting distribution shown in Table 7.

Table 7

Partial Distribution of Operating Gain

	Enterprise Operating Gain 1998
5% Perc =	(71,619,600)
10% Perc =	(50,000,000)
15% Perc =	(32,521,600)
20% Perc =	(10,258,950)
25% Perc =	(763,150)
30% Perc =	12,859,600
35% Perc =	38,245,700
40% Perc =	60,052,150
45% Perc =	84,517,300

This is a portion of the full distribution that increases in 5% increments up to 95%. This table communicates for example that 5% of the operating gain samples are less than $(71,619,600) and that 30% of the samples are less than $12,859,600. At a 90% confidence level $(50,000,000) is the operating gain. Similarly at a 95% confidence level the gain is $(71,619,000).

At the 90% confidence level, surplus would be decreasing at most $50,000,000 in the year. If the company started with only $50,000,000 this would not be a pleasant outcome. So a second constraint is necessary, that is 'what is the maximum proportion of surplus the company management is willing to lose in any year?' For example, management is willing to risk a decrease to be at most 10% of starting surplus. In other words, the $(50,000,000) would equal (0.10) times the needed surplus. The surplus needed would be calculated as follows:

$(50,000,000) / (0.10) = $500,000,000 surplus needed

The calculated surplus is the theoretical amount needed to support business as a going-concern under the stated constraints. This amount should be used in the ROE and pricing calculations. Comparing this surplus to the enterprise surplus may indicate a redundancy or deficiency. If the calculated surplus is less than the company surplus, the redundancy isn't necessarily excess to squander. The total enterprise surplus may need to be maintained for statutory or regulatory purposes.

RUIN WITH ROE MEASURES

Many insurance companies are being evaluated from a financial viewpoint. The question that comes up is the level of ROE that the company wants to target. The level of surplus affects this ROE measure. Lower surplus translates into a higher leverage ratio increasing the potential ROE while generating a greater chance of ruin. To reduce the chance of ruin more surplus would be held, reducing the ROE. This puts the insurance company in a precarious position.

With the DFA model it is possible to test out different levels of surplus. One can begin with a certain level of surplus, capturing the appropriate values for ROE and ending surplus. Different levels of surplus translate into differing ROE averages. Accompanying each ROE average is a probability of ruin distribution. An optimization then has to be made on the risk and return trade off.

CONCLUSION

As the line between the financial industry and the insurance industry blurs, actuaries are becoming the financial leaders in the insurance industry. From a financial perspective there is a strong desire to allocate surplus to measure, track, and rate performance on a line of business basis. There are many ways to allocate surplus once the overall needed surplus amount is determined. Of the methods presented, the variation method incorporates the most characteristics desired from a surplus allocation method. However, this is just the starting point for others to build upon and to improve.

Distribution by Reserves

	(1) Loss Reserves	(2) Unearned Premiums	(3) Sum	(4) Distribution	(5) Estimated Surplus
Homeowners	66,900,470	27,277,000	94,177,470	7.3%	36,619,972
P Auto Liab	385,914,100	44,801,000	430,715,100	33.5%	167,479,280
P Auto Phys Dam	37,426,200	41,044,000	78,470,200	6.1%	30,512,356
C Auto Liab	112,318,300	6,093,000	118,411,300	9.2%	46,043,055
C Auto Phys Dam	3,599,136	2,218,000	5,817,136	0.5%	2,261,935
CPP Liab	141,808,400	51,320,000	193,128,400	15.0%	75,096,056
CPP Prop	63,105,810	68,577,000	131,682,810	10.2%	51,203,550
Other Liab	3,725,144	7,565,000	11,290,144	0.9%	4,390,060
Other Liab - Umbrella	1,394,024	316,000	1,710,024	0.1%	664,926
Workers Comp	146,415,200	74,058,000	220,473,200	17.2%	85,778,809
Personal	490,240,770	113,122,000	603,362,770	46.9%	234,611,608
Commercial	472,366,014	210,147,000	682,513,014	53.1%	265,388,392
Enterprise	962,606,784	323,269,000	1,285,875,784	100.0%	500,000,000

(1) Year end 1997 Net Loss Reserves
(2) Year end 1997 Unearned Premium Reserves
(4) (3) / Enterprise (3)
(5) (4) * Enterprise (5)

Exhibit 2

DURATION CALCULATION

Line of Business: Homeowners

Accident Years

Calendar Year	(1) 1988	(2) 1989	(3) 1990	(4) 1991	(5) 1992	(6) 1993	(7) 1994	(8) 1995	(9) 1996	(10) 1997	(11) Total	(12) Year	(13) Present Value 6.53%	(14) Weighted PV
1997	2,671	1,005,428	1,248,769	1,235,685	1,468,734	4,554,592	9,874,726	10,488,831	31,928,154	48,079,825	109,887,415	0.5	106,466,259	53,233,129
1998		1,071,315	2,672,012	1,972,956	1,901,699	4,189,929	7,045,897	7,282,894	35,960,250	83,542,241	145,639,193	1.5	132,455,615	198,683,422
1999		983,868	36,768	652,113	4,388,283	1,902,007	5,579,526	2,659,516	15,570,392	37,330,497	69,102,970	2.5	58,995,230	147,488,074
2000			36,833	996,306	716,188	2,585,618	4,265,356	1,372,445	8,610,860	12,484,401	31,068,007	3.5	24,897,839	87,142,436
2001				306,002	104,284	551,020	951,525	2,545,685	2,007,813	6,180,550	12,646,879	4.5	9,513,924	42,812,659
2002					102,661	739,430	585,785	1,066,333	1,142,832	1,595,912	5,232,953	5.5	3,695,313	20,324,222
2003						1,268,525	26,357	2,154,525	1,437,799	1,631,019	6,518,225	6.5	4,320,777	28,085,049
2004							26,357	5,142,536	43,174	308,906	5,520,973	7.5	3,435,391	25,765,432
2005								2,153,695	21,587	576,838	2,752,120	8.5	1,607,519	13,663,907
2006								1,245,852	21,587	55,009	1,322,448	9.5	725,096	6,888,409
2007								784,585		55,009	839,594	10.5	432,130	4,537,362
2008								532,565			532,565	11.5	257,303	2,958,989
2009								235,252			235,252	12.5	106,693	1,333,657
2010											0	13.5	0	0
2011											0	14.5	0	0
2012											0	15.5	0	0
2013											0	16.5	0	0
2014											0	17.5	0	0
2015											0	18.5	0	0
2016											0	19.5	0	0
2017											0	20.5	0	0
2018											0	21.5	0	0
2019											0	22.5	0	0
2020											0	23.5	0	0
2021											0	24.5	0	0
													346,909,087	632,916,747

(15) Macaulay Duration 1.8244

(16) Modified Duration 1.7126

(1) - (10) Calculated in the DFA
(11) Sum of columns (1) through (10)
(12) Year of payment assuming uniform over a given year
(13) (11) / [(1 + interest rate) ^ (12)]
(14) (13) * (12)
(15) Total (14) / Total (13)
(16) (15) / (1 + interest rate)

Exhibit 3

Duration Distribution

Duration	Duration	Relativity	(S/P)	Premium	Estimated Surplus	Split
	(1)	(2)	(3)	(4)	(5)	(6)
Homeowners	1.8698	0.8023	0.4814	54,552,830	26,261,991	5.4%
P Auto Liab	2.0946	0.8988	0.5393	179,204,200	96,644,215	19.7%
P Auto Phys Dam	1.2064	0.5177	0.3106	164,174,860	50,994,823	10.4%
C Auto Liab	2.5792	1.1068	0.6641	24,369,986	16,183,090	3.3%
C Auto Phys Dam	1.4776	0.6340	0.3804	8,872,202	3,375,212	0.7%
CPP Liab	3.2283	1.3853	0.8312	102,639,868	85,311,534	17.4%
CPP Prop	1.8037	0.7740	0.4644	137,153,660	63,692,602	13.0%
Other Liab	2.6794	1.1498	0.6899	15,129,588	10,437,285	2.1%
Other Liab - Umbrella	2.6639	1.1431	0.6859	630,543	432,473	0.1%
Workers Comp	3.6035	1.5463	0.9278	148,116,140	137,419,704	28.0%
Personal	1.8861	0.8093		397,931,890	173,901,029	35.4%
Commercial	2.7944	1.1991		436,911,987	316,851,898	64.6%
Enterprise	2.3304	1.0000		834,843,877	490,752,928	

	Adjusted Surplus	P/S
	(7)	(8)
Homeowners	26,756,836	2.0388
P Auto Liab	98,465,245	1.8200
P Auto Phys Dam	51,955,699	3.1599
C Auto Liab	16,488,021	1.4780
C Auto Phys Dam	3,438,810	2.5800
CPP Liab	86,919,027	1.1809
CPP Prop	64,892,737	2.1135
Other Liab	10,633,951	1.4228
Other Liab - Umbrella	440,622	1.4310
Workers Comp	140,009,051	1.0579
Personal	177,177,781	2.2459
Commercial	322,822,219	1.3534
Enterprise	500,000,000	1.6697

(1) Duration
(2) Line Duration / Enterprise Duration
(3) (2) * Surplus/Premium ratio of 0.60 = (1/1.6697)
(4) Premium
(5) (3) * (4) = (Surplus/Premium) * (Premium) = Estimated Surplus
(6) (5) / (Enterprise 5)
(7) (6) * Enterprise
By Line: [(5) / Consumer or Commerical (5)] * Consumer or Commercial (7)
(8) (4) / (7)

Variation Distribution

Operating Gain per $ WP Standard Deviation	(1) 1998	(2) 1999	(3) 2000	(4) 2001	(5) 2002	(13)	(6) Expected Value	(7) Within Variance	(10) K factors
Home	0.4532	2.2769	0.6970	0.6594	0.4550		0.9083	0.4784	1.5015
P Auto Liab	0.1574	0.1235	0.1461	0.1395	0.1536		0.1440	0.0001	0.0004
P Auto Phys Dam	0.1199	0.9526	0.2524	0.2326	0.1184		0.3352	0.0984	0.3088
C Auto Liab	1.3752	1.4673	1.6137	1.6697	1.7350		1.5722	0.0175	0.0550
C Auto Phys Dam	0.4705	2.3967	0.7265	0.6692	0.5526		0.9671	0.5195	1.6302
CPP Liab	0.0575	0.0646	0.1012	0.0772	0.0641		0.0769	0.0002	0.0007
CPP Prop	0.5931	4.5416	1.1913	1.0898	0.5455		1.5923	2.2411	7.0333
Other Liab	0.1688	0.2055	0.2183	0.2339	0.2391		0.2131	0.0006	0.0020
Other Liab - Umbrella	0.8574	0.9288	1.1948	1.2197	1.3292		1.1060	0.0328	0.1028
Workers Comp	0.1308	0.1456	0.1748	0.1856	0.2019		0.1677	0.0007	0.0021
Personal	0.1255	0.7125	0.2001	0.1830	0.1159				
Commercial	0.2088	1.4789	0.3995	0.3605	0.2009				
Enterprise	0.1544	1.1122	0.2955	0.2666	0.1367	0.3931			
							(8)		
Expected Value	0.4384	1.3103	0.6316	0.6197	0.5415		0.7063		
							(9)		
Between Variance	0.1571	1.8391	0.2649	0.2702	0.2798		0.3186		

Credibility Factors	(11)	Operating Gain per $ WP Credibility Weighted Standard Deviations	(12)
Home	0.7691	Home	0.8621
P Auto Liab	0.9999	P Auto Liab	0.1441
P Auto Phys Dam	0.9418	P Auto Phys Dam	0.3569
C Auto Liab	0.9891	C Auto Liab	1.5628
C Auto Phys Dam	0.7541	C Auto Phys Dam	0.9035
CPP Liab	0.9999	CPP Liab	0.0770
CPP Prop	0.4155	CPP Prop	1.0756
Other Liab	0.9996	Other Liab	0.2133
Other Liab - Umbrella	0.9796	Other Liab - Umbrella	1.0980
Workers Comp	0.9996	Workers Comp	0.1680

Operating Gain per $ WP Credibility Weighted Standard Deviation/ Standard Deviation of Enterprise	(14)	(Std Dev ratio) * (Surplus/Prem ratio) (15)		Premium (16)
Home	2.1932	1.3133		54,552,816
P Auto Liab	0.3665	0.2195		179,199,580
P Auto Phys Dam	0.9079	0.5436		164,176,060
C Auto Liab	3.9757	2.3807		24,369,966
C Auto Phys Dam	2.2985	1.3763		8,872,197
CPP Liab	0.1959	0.1173		102,639,824
CPP Prop	2.7363	1.6385		137,153,600
Other Liab	0.5427	0.3250		15,129,588
Other Liab - Umbrella	2.7933	1.6727		630,538
Workers Comp	0.4273	0.2559		148,116,220
			Consumer	397,928,456
Premium/Surplus ratio	1.67		Commercial	436,911,953
			Enterprise	834,840,409
Surplus/Premium ratio	0.60 (assumed)			

	Estimated Surplus (17)	(18)	(19)	Surplus (19)	P/S (20)
Home	71,644,576	13.0%		65,002,064	0.8392
P Auto Liab	39,331,270	7.1%		35,684,679	5.0218
P Auto Phys Dam	89,251,941	16.2%		80,976,965	2.0274
C Auto Liab	58,017,003	10.5%		52,637,969	0.4630
C Auto Phys Dam	12,211,073	2.2%		11,078,926	0.8008
CPP Liab	12,040,677	2.2%		10,924,328	9.3955
CPP Prop	224,728,029	40.8%		203,892,415	0.6727
Other Liab	4,916,523	0.9%		4,460,688	3.3918
Other Liab - Umbrella	1,054,673	0.2%		956,869	0.6589
Workers Comp	37,898,862	6.9%		34,385,077	4.3078
Consumer	200,227,787	36.33%	181,663,708		2.1905
Commercial	350,866,839	63.67%	318,336,292		1.3725
Enterprise	551,094,626		500,000,000		1.6697

Notes:
(1) - (5) Standard Deviations of Operating gain per Dollar of Written Premium
(6) - Mean of years by line of business
(7) - Variance of years by line of business
(8) - Mean of (6)
(9) - Variance of Expected Values (6)
(10) - (7) / (9)
(11) - n / (n + K) = Z where n = # of years
(12) - (6) * (11) + (8) * (1 - (11))
(13) - Mean of Enterprise
(14) - (12) / (13)
Distribution of capital
(15) - (14) * selected surplus/premium ratio
(16) - Year end 1997 Premium
(17) - (15) * (16)
(18) - (17) / Enterprise (17)
(19) - (18) * Enterprise (19)
(20) - (16) / (19)

DFA MODEL USED

The following includes excerpts from the papers "Building a Public Access PC-Based DFA Model" (1997 CAS Summer Forum, Vol.2) [2] and "Using the Public Access DFA Model: A case Study" (1998 CAS Summer Forum) [3]. Both papers are used with the permission of the papers' authors.

The Dynamic Financial Analysis (DFA) model used in this paper is a public access model. **The actuarial consulting firm Miller, Rapp, Herbers & Terry (MRH&T) created Dynamo2**. **Dynamo2** is Excel based enabling the user to create calculations as needed.

Each iteration of the model starts with detailed underwriting and financial data showing the historical and current positions of the company. It randomly selects values for 4,387 stochastic variables, calculates the effect on the company of each of these selected values, and produces summary financial statements of the company for the next five years based on the combined effect of the random variables and other deterministic factors.

The model consists of several different modules, each of which calculates a component of the model indications. Separate modules are included for investments, catastrophes, underwriting, taxation, interest rates, and loss reserve development. The number of lines of business can be expanded or contracted to fit the needs of the user. The model used allows for ten different lines of business:

- Homeowners
- Private Passenger Auto Liability
- Private Passenger Auto Phys. Dam.
- Commercial Auto Liability
- Commercial Auto Phys. Dam.
- CPP - Liability
- CPP - Property
- Other Liability
- Umbrella
- Workers Compensation

For each line of business, the underwriting gain or loss is calculated separately for: 1) new business, 2) 1st renewal business and 3) 2nd and subsequent renewals. This division is provided to reflect the aging phenomenon, in which loss experience improves with the length of time a policyholder has been with a company. These three categories are then added to calculate underwriting results on a direct, ceded, and net basis.

The values for each simulation are shared among the different modules. Thus, if the random number generator produces a high value for the short term interest rate, this high interest rate is used in the investment module as well as the underwriting module. Similarly, a high value for catastrophes in the catastrophe module carries through to the reinsurance and underwriting modules.

The primary risks that are reflected in the model are:

- Pricing risk
- Loss reserve development risk
- Catastrpohe risk
- Investment risk

CAVEATS OF THE MODEL

Some factors, having a potentially significant impact on results, are omitted from the model because, in the opinion of the authors, they are beyond the scope of an actuarial analysis. For example, fraud by managers is a leading cause of insurance insolvency. Whether fraud is likely to occur (or is currently occurring) at a particular insurer, is not something an actuary is qualified to ascertain. Thus, any financial effects from fraudulent behavior are simply omitted from the model. Other examples of omitted factors that definitely could have a significant effect on insurance operations include a change in the tax code, repeal of the McCarran-Ferguson Act, a major shift in the application of a legal doctrine or the risk of a line of business being socialized by a state, province, or federal government. Thus, the range of possible outcomes from operating an insurance company is actually greater than a DFA model would indicate; the model is designed to account only for the risks that can be realistically quantified.

The values used as input in the model are derived from past experience and current operational plans. To the extent that something happens in the future that is completely out of line with past events, the model will be inaccurate. For example, the size of a specific catastrophe is based on a lognormal distribution with the parameter values based on experience over the period 1949 - 1995 (adjusted for inflation). However, if this process had been used just prior to 1992, the chance of two events occurring within the next 2 ½ years, both of which exceeded the largest previous loss by a factor of more than 2, would have been extremely small. However, Hurricane Andrew caused $15.5 billion in losses in August of 1992 and the Northridge earthquake caused $12.5 billion in insured losses in January 1994. The largest insured loss prior to that was Hurricane Hugo, which had caused $4.2 billion in losses in 1989. Also, if changes in any operations occur, then the results would not be valid.

The DFA model encompasses catastrophes, which have a significant impact on the property lines of business. The liability lines of business are more influenced by changes in public attitudes, and legislative or judicial changes. These changes are difficult if not impossible to model accurately. The variation method considers these to the extent that they are captured in historical data and variations.

The number of years used may affect the credibility of results. The DFA model results have a compounding effect from year to year (e.g., the first year results are used in the second year, the second year results are used in the third year, and so on). With nominal growth assumptions, this will result in larger variation for the more distant years. If ample simulations are run, then the distant years' variation becomes more stable.

When a significant legislative or judicial change occurs, the model should be adjusted to reflect such changes. The surplus allocation process should be run once again to incorporate these changes.

MODEL USAGE

Before relying on a DFA model for **any** purpose the user must be comfortable with the inputs and the outputs. This includes using it to allocate surplus.

Duration Adjustments to the DFA Model

The assumption was made that all payments would be made by the end of the twenty-first year for each accident year, however in the original model only five calendar years of payments were calculated for each accident year. These payments needed to be extended to twenty years past the last projected accident year. Extended payments were produced in the same fashion as done by the model for the first five years.

After projecting how much is going to be paid from each accident year for any calendar year, these payments are summed across all accident years for the appropriate calendar year. This generates projected calendar year paid amounts as in column 1 of Table 2. The total column in Exhibit 2 also shows calendar year paid amounts.

A discount rate was needed to find the present value of these calculations. The discount rate used came from the first year's projected investment information of the model. Dividends, coupon payments, and interest were summed and divided by the average book value amount invested in stocks, bonds, and cash over the year. This avoids both realized and unrealized capital gains or losses. By calculating a DFA discount rate, it allows the interest rate to vary with the projected economic conditions for each iteration.

For calculating present values, a uniform payment pattern during each year was assumed, giving an average payment mid-way through the year.

Investment Return Adjustments to the DFA Model

To capture interest earned by line of business, adjustments had to be made to the DFA model.

The pay-out rate of reserves was determined from the payment patterns already mentioned. The percentage of reserves available for investment over the course of the year is:

$$[\ \{ (1 - PP_T) + (1 - PP_{T+1}) \} / 2\ ] / (1 - PP_T)$$

Where: PP = Payment Pattern (Cumulative % Paid)
T = Beginning of Calendar Year
T+1 = End of Calendar Year

The division by 2 in the formula assumes the payment of reserves is made uniformly over the calendar year.

This calculation is done for each Calendar Year / Accident Year combination needed.

Using the above calculation, the amount available for investment was found by multiplying the percentage of reserves available by the appropriate reserves.

$$[\ [\ \{ (1 - PP_T) + (1 - PP_{T+1}) \} / 2\ ] / (1 - PP_T)\ ] * \text{Reserves}$$

Where : When Calendar Year = Accident Year
Reserves = Reserves at the end of the Calendar Year

Otherwise
Reserves = Reserves at the beginning of the Calendar Year

Example: For a given line of business and accident year, 20% of the losses had been paid by the beginning of the current calendar year and 40% paid by the end. The year began with $5,000,000 in reserves for this particular line and accident year.

The amount available for investment is [{ (1-.20) + (1-.40) } /2]/(1-.20) = 0.875. In other words, 87.5% of the beginning of the year reserves are available for investment over the course of the year, or $5,000,000 * .875 = $4,375,000.

The method of calculating the rate of return used for this method was based on the market value return. The ending market value of the stocks, bonds, and cash were added to the sum of dividends, coupon payments, and interest received. The resulting amount was divided by the sum of the beginning market value of the stocks, bonds, and cash. A different rate of return was determined for each calendar year.

Credibility Weightings

In the Variation Method

Even after running 1,000 iterations, the information is not necessarily fully credible since this calculation deals with the standard deviation. The model itself should be fully credible, but the standard deviation deals with the number of samplings. If enough iterations are run, the standard deviations should be relatively stable from year to year. Due to computing limitations, only 1,000 iterations were run, which lacks full credibility.

Table 8 below lays out the credibility weighting of the standard deviations. Applying the Bühlmann credibility across the years and between the lines with the use of columns 6 through 10, credibility is determined by line of business and displayed in column 11. This process is shown explicitly in Exhibit 4. Giving credibility weight to the expected value for a line of business over the years (column 6) and the complement to the average of the expected values for all lines of business (column 8) results in a credibility weighted standard deviation of net operating gain per dollar of net written premium (column 12). This is the main factor in helping determine the distribution of surplus. The rest of the steps are similar to those used in the duration method. The Bühlmann credibility is further described in Appendix E.

Table 8

Standard Deviation of Net Operating Gain Per Dollar of Net Written Premium

	(1) 1998	(2) 1999	(3) 2000	(4) 2001	(5) 2002	(6) Expected Value	(7) Within Variance	(10) K Factors
Home	0.4532	2.2769	0.6970	0.6594	0.4550	0.9083	0.4784	1.5015
P Auto Liab	0.1574	0.1235	0.1461	0.1395	0.1536	0.1440	0.0001	0.0004
P Auto Phys Dam	0.1199	0.9526	0.2524	0.2326	0.1184	0.3352	0.0984	0.3088
C Auto Liab	1.3752	1.4673	1.6137	1.6697	1.7350	1.5722	0.0175	0.0550
C Auto Phys Dam	0.4705	2.3967	0.7265	0.6892	0.5526	0.9671	0.5195	1.6302
CPP Liab	0.0575	0.0646	0.1012	0.0772	0.0841	0.0769	0.0002	0.0007
CPP Prop	0.5931	4.5416	1.1913	1.0898	0.5455	1.5923	2.2411	7.0333
Other Liab	0.1688	0.2055	0.2183	0.2339	0.2391	0.2131	0.0006	0.0020
Other Liab – Umbrella	0.8574	0.9288	1.1948	1.2197	1.3292	1.1060	0.0328	0.1028
Workers Comp	0.1308	0.1456	0.1748	0.1856	0.2019	0.1677	0.0007	0.0021
Personal	0.1255	0.7125	0.2001	0.1830	0.1159			
Commercial	0.2088	1.4789	0.3995	0.3605	0.2009		(13)	
Enterprise	0.1544	1.1122	0.2955	0.2666	0.1367		0.3931	
						(8)		
Expected Value [4]	0.4384	1.3103	0.6316	0.6197	0.5415	0.7083		
						(9)		
Between Variance [5]	0.1571	1.8391	0.2649	0.2702	0.2798	0.3186		

(6) = average of (1) to (5)

	(11) Credibility	(12) Cred. Wtd. Std Deviations
Home	0.7691	0.8621
P Auto Liab	0.9999	0.1441
P Auto Phys Dam	0.9418	0.3569
P Auto Liab	0.9891	1.5628
C Auto Phys Dam	0.7541	0.9035
CPP Liab	0.9999	0.0770
CPP Prop	0.4155	1.0756
Other Liab	0.9996	0.2133
Other Liab – Umbrella	0.9798	1.0980
Workers Comp	0.9996	0.1680

[4] The expected value is a straight average of the individual line of business data points.

[5] The between variance is the sum of the squared differences between the line of business data point and the expected value all divided by the number of lines of business.

BÜHLMANN CREDIBILITY

Bühlmann credibility is based on the formula n / (n + K) = Z

Where: n is the number of observations

K is the within variance / between variance

Z is the crebibility factor

The within variance is calculated within the same class or line of business across years or periods. In this application it would be the variance for a certain line of business over the 5 year period.

$$\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

Xi is an individual observation

$\bar{X}$ is the average observation within the line of business

n is the 5 years

The between variance is calculated within the same year but across lines of business.

$$\frac{\sum_{i=1}^{m} (Y_i - \bar{Y})^2}{m}$$

Yi is an individual observation in the year

$\bar{Y}$ is the average observation for the year over all lines of business

m is the number of lines of business which is equal to 10 in our application

REFERENCES

[1] Calfarm v. Deukmejian, 48 Cal. 3d 805, 771 P.2nd 1247 (1989).

[2] D'Arcy, S. P., R. W. Gorvett, J. A. Herbers, T. E. Hettinger, S. G. Lehmann, and M. J. Miller, 1997, "Building a Public Access PC-Based DFA Model," *Casualty Actuarial Society Forum,* Fall 1997, Vol. 2, pp. 1-40.

[3] D'Arcy, S. P., R. W. Gorvett, T. E. Hettinger, and R. J. Walling, III, 1998, "Using the Public Access DFA Model: A Case Study," *Casualty Actuarial Society Forum,* Summer 1998, pp. 53-118.

[4] Ettlinger, K.H.; Hamilton, K.L.; and Krohm,G., *State Insurance Regulation* (First Edition), Insurance Institute of America, 1995, Chapter 8, pp. 209-231.

[5] Feldblum, S. "Pricing Insurance Policies: The Internal Rate of Return Model," May 1992.

[6] Ferguson, R. E., "Duration," *Proceedings of the Casualty Actuarial Society*, Vol. LXX, 1983, pp. 265-288.

[7] Miller, M. and Rapp, J., "An Evaluation of Surplus Methods underlying Risk Based Capital calculations," *1992 Discussion Paper Program,* Volume 1, pp. 1-122.

[8] Herzog, T. N., "Bühlmann's Approach," *Introduction to Credibility Theory* (Second Edition), 1996, pp. 69-98.

[9] Stone, J.M., "A Theory of Capacity and the Insurance of Catastrophe Risks," *The Journal of Risk and Insurance*, June 1973, Vol. XL No. 2, Part I, pp. 231-243, and September 1973, Vol. XL No. 3, Part II, PP. 339-355.