The Casualty Actuarial Society *Forum*

Summer 2003 Edition

Including the 2003
Enterprise Risk Management & Dynamic Finanical Analysis Modeling
Call Papers

To CAS Members:

This is the Summer 2003 Edition of the Casualty Actuarial Society *Forum*. It contains an overview of enterprise risk management, three ERM & DFA Modeling Call Papers, and six additional papers.

The *Forum* is a nonrefereed journal printed by the Casualty Actuarial Society. The CAS is not responsible for statements or opinions expressed in the papers in this publication. See the Guide for Submission to CAS *Forum* (www.casact.org/aboutcas/forum.htm) for more information about the *Forum*.

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,

Joseph A. Smalley

Joseph A. Smalley, CAS *Forum* Chairperson

The 2003 CAS Enterprise Risk Management & Dynamic Financial Analysis Modeling Call Papers Presented at the 2003 Risk and Capital Management Seminar July 28-29, 2003 Capital Hilton Washington, D.C.

The Summer 2003 Edition of the CAS *Forum* is a cooperative effort between the CAS *Forum* Committee, the CAS Committee on Dynamic Risk Modeling (formerly the Dynamic Financial Analysis Committee), and the Enterprise Risk Management Committee.

The CAS Committee on Dynamic Risk Modeling and the Enterprise Risk Management Committee present for discussion three papers prepared in response to their 2003 call for papers, entitled "ERM & DFA Modeling: Risk Correlation, Integration, Dependency, and Concentration."

This *Forum* includes papers that will be discussed by the authors at the 2003 CAS Risk and Capital Management Seminar, July 28-29, 2003, in Washington, D.C.

ERM & DFA Modeling Call Papers

Committee Report

Additional Papers

ERM and DFA Using Active Knowledge Structures

Jim Brander and Sam Manoff

ERM and DFA using Active Knowledge Structures

Jim Brander
Interactive Engineering
Email: jbrander@bigpond.net.au

Sam Manoff
Tupai Business Systems
Email: smanoff@tupaisystems.com

Abstract

Active, uncommitted knowledge structures are described as a means of representing risks with either a stochastic history or a hypothetical value, or a mixture of the two within a single risk. These same knowledge structures are shown to contribute to an increased speed of organizational response in a dynamic environment and to the ability of a simulation model to learn from its own operation.

Introduction

This paper addresses the following areas:

- **Correlation/dependency:** The storing of correlations and dependencies
- **Integration:** Methodology for integrating correlated risk distributions into models
- **Dependency/causal models:** The modeling of risk defined by presumed causes

Lying behind these areas of interest is an increasing interconnectedness and dynamicity of the risk environment – conditions which current analytic tools do not well support. Tools now implement one of several techniques - the 'natural order' of calculation of spreadsheets (which must be known in advance of any calculation), directed dataflow or specific programming using a single point of control. None of these techniques is appropriate for interdependent risks, where the nature of the interaction is dynamic.

We will attempt to show that an undirected structure with distributed control and comparatively complex messaging, and with the abilities to store experiential knowledge within its structure and to modify its own structure, addresses the areas mentioned above. The resulting model looks very much like current analytic models – it is just that the underlying process that propagates information is very different.

Restriction on Information Transmission leads to a Disconnect

A model is intended to assist in analysing a complex situation. The representation of enterprise risk, involving as it does probability and the connections among different risks, is a complex problem, modeled only with difficulty. Unfortunately, much of the current modeling effort is disconnected from the actual problem, and instead turns on artifacts of the analytic modeling process, as the following diagram seeks to illustrate.

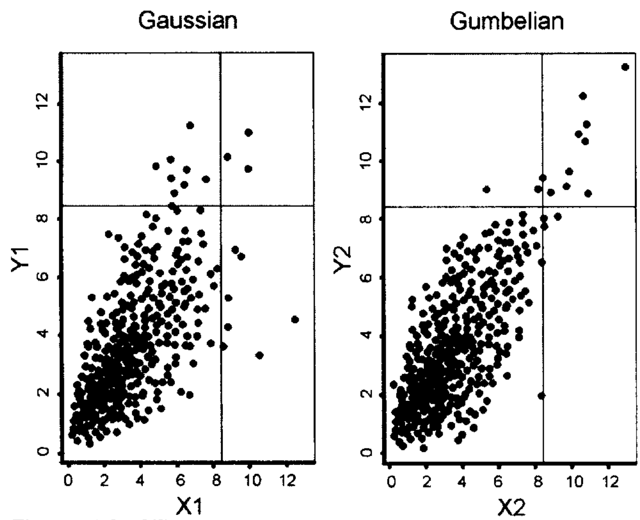


Figure 1 (after [5])

Deciding whether the association between two risks is Gaussian or Gumbelian would seem to be far from the actual problem. Large events are rare, so a few more events can quickly invalidate any a priori modeling decision. The other aspect with risk events in extremis is that the extreme events are likely to involve size or resource thresholds and thus processes that simply do not apply in less extreme events – a large, critically situated, earthquake could destroy stockmarkets around the world, whereas there is no correlation with a small tremor. That is, large risks are much more likely to be interconnected because the large-scale processes they unleash will overlap. The current methods of closed form analysis handle interdependence poorly. This may be one reason for the popularity of DFA using simulation, as the cyclic nature of simulation appears to allow interdependence to be handled with minimal effort. This is somewhat of an illusion, particularly when a risk is spiked as part of a simulation.

We begin by examining whether the statistical methods of current risk modelling methods are appropriate or necessary

When the only object in an analytic model that can be propagated is a number or a group of numbers, description of the risk must be turned into, at most, a few numbers that act as parameters – a type of distribution, a variance, possibly a skew, a correlation, a copula. This restriction, whether for insurance or finance or engineering, is artificial – it declares a clarity of parameter and an initial precision that does not exist in the messy world we are attempting to model. The actuarial literature has many examples where the author points out that the distributions involved are far from Gaussian, and then proceeds to use the method anyway, for want of something better.

Beginning a calculation with a number asserts that there is a seed that can be known precisely. This is valid for a payroll application, where the rate per hour is known for a particular employee and operations on this seed will yield a valid result – it is not valid where a calculation must start with multiple risks which are interdependent and known only by their probability distributions, or where the end result is non-monotonic – an example is interest rates, where very low and very high values both lead to asset inflation. Some dependencies, such as an earthquake causing fires, are directional, allowing a simple directed approach. Whether in their language or in their mental models, humans do not rely on precise numbers to begin their processing, but rather on influences and associations, although the end result of the mental activity can still have astounding precision. If we broaden what we can propagate in our analytic models to allow for the propagation of numeric ranges and the storage of distributions and stochastic associations, then many of the limitations of current analytic modeling techniques using singular values disappear.

Numbers in analytic models are typically represented by values in memory that can be loaded into a register at the behest of a procedure, and there manipulated. Let us instead use a network object to represent a number – something with existence, attributes and the ability to be linked. As a very brief introduction to knowledge networks, we offer the equation shown in Figure 2.

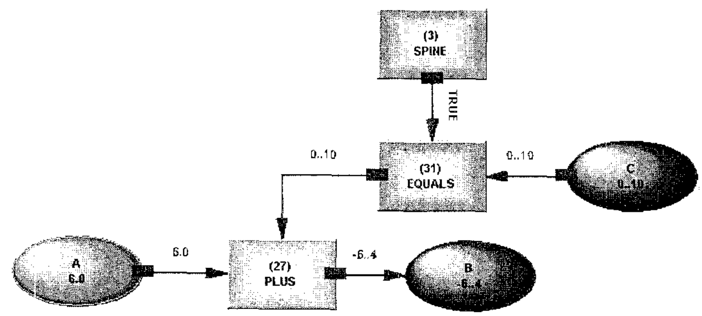


Figure 2 – A + B = C

The structure is made up of variables, operators (PLUS etc.) and links. As shown, the variable A has a singular value and C has a range, and these two values have been combined at the PLUS operator to produce a value which is propagated to B – the direction of information flow was dynamically determined [2]. The SPINE operator, in the top center of the diagram, functions as an AND, connecting and controlling all the statements in the model, allowing the model to become a controllable submodel in a larger model if desired. A True logical state from the SPINE has enabled the EQUALS operator to propagate the value from C. As states and values of the variables come and go, the direction of information transmission in the structure may change, reflecting the current state (direction of propagation is part of that state). The undirected nature of the structure means that it is initially uncommitted as to purpose – useful when dealing with situations such as interdependent risk, where the influence can come from any direction. Operators only perform calculations when the logical states on their connections change, so that the network micro-schedules its activity, and no external control algorithm is required – a desirable feature as complexity grows. To elaborate a little more, we may have the statement

IF A + B = C THEN D + E = F

The numeric elements embedded in the logical statement look like the a-b-c structure we have already encountered. In the interests of generality, it would seem reasonable to use identical structure in both places in the statement. If we implement the logical part of the statement as sentential logic, we have the arrangement shown in Figure 3.

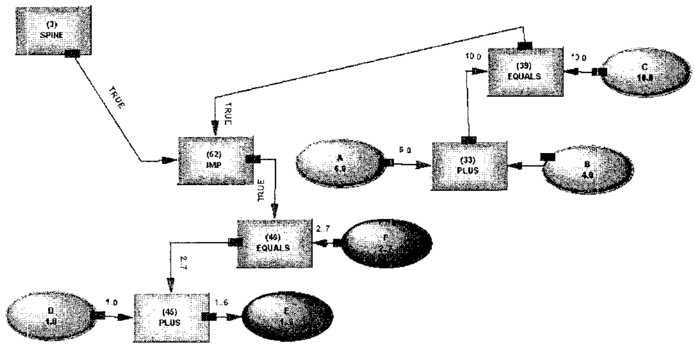


Figure 3 – IF A + B = C THEN D + E = F

The IF...THEN... is implemented as a logical implication, and the logical operator also responds to changes in logical states on its connections. The example shows structures as components in a larger structure, and we have retained all the possible inferences - modus tollens, for example. This allows the system to reason about the cause when an effect is not as expected.

Without going into details, there are structural analogs of the usual programming machinery - FOR and WHILE loops (see [3]), fetch and store operations, sigma and other analytic operators used in actuarial models, string, list and object handling – all involving identifiable states in structure.

The foregoing may give the impression of merely a graphical representation of textual statements, but that would be a false impression. As relations become dense and complex, graphical representations except at a high level of abstraction are much harder to comprehend than textual representations– the many-variabled relationship between earthquake magnitude and damage shown in [1], for example. Instead, the text of an analytic statement is converted into a structure, a structure which stores its current state inside itself and propagates states and values through its connections, and which interacts with the structure produced by other statements. What distinguishes this structure from a graph is its ability to alter its connections – to form what appears to be a new graph while the structure (the totality of states and connections) remains invariant. This may seem an unusual occurrence, so

X = SUM(List)

will serve as a common example. Some calculation is involved in determining the list, and we don't yet know if we are working out the value of X or some member of the list, based on X and other values we find, and we need to recover our original state if the value for the list is lost. The underlying paradigm is of an active, undirected, extensible and self-modifying structure, rather than statements generating a sequence of instructions or a graph that is "understood" and operated on by an external algorithm. The knowledge network structure can appear very similar to the dataflow paradigm, where inputs flow to an operator, which then produces an output, which flows to another operator as an input, and so on until calculation is complete. The dataflow paradigm, however, assumes that the flow path of the information can be predetermined, directions remain fixed and the declared topology is invariant.

More Complex Messages

To use numeric ranges effectively, we actually need to go one step back in how our models work. If we have a number that is calculated in a model, exactly when can we access that number. Obviously, when the calculation is finished – but the calculation may be waiting on another calculation, and so on. And we may be using the statement as a statement rather than procedurally – that is, if we have

A = B + C

we may be working out C based on values for A and B. At this point, someone may object that "But I only want to work out A, not anything else". If we wish to work out interdependent risks, and the interdependence needs to be dynamically determined, then we should allow the model to determine the situation, not hope that we can program it in advance. If we associate a state with the number, the state tells us whether the number is valid. If we are to use a state, it can't be a Boolean state – we may not yet have found the number but are still looking for it, or perhaps we failed to find it, or we may have encountered an error in the process. We already have True and False when handling logical variables, so let us use False as the state to indicate that we have a numeric range rather than a single value. The numeric range is also an object (itself comprised of objects), so it can be

An integer range 1..10
A real range 2.35<->7.9

The range does not have to be contiguous, so -3..5, 7..21, 43..1000000 is acceptable. These range objects are dynamically constructed and propagated, so the limits of the range do not have to be known beforehand. The Modus Tollens inference we mentioned has value in a statement like IF A < B THEN C > D, where influences flow in any direction rather than a test causes an action.

By allowing numeric ranges on variables to interact and cut each other in the manner of Constraint Reasoning – that is, information can flow back and forth on a connection - a continually reducing solution space is obtained. This reducing space, driven by many interacting influences, permits the solution of interdependent risk in closed form analysis.

Distributions and Means of Correlation

But ranges aren't enough – for problems involving uncertainty, we need to represent probability over a range – a distribution. A discrete distribution is clunky if we can only use numbers to represent the different bins – the increment between the bins needs to be a preset constant. Now that ranges are available to be used for the bin limits, the ranges can be adaptive, with the bins wide where hits (occurrences) are few, narrow for precision where hits are many, and nonexistent where there are no hits (that is, the range for each bin is contiguous, but the ranges need not be contiguous). Figure 4 shows a simple example of a distribution (the value on the Y axis is number of occurrences within the range at its foot, making it an occurrence distribution rather than a probability distribution, but it is easily converted, based on the current range):

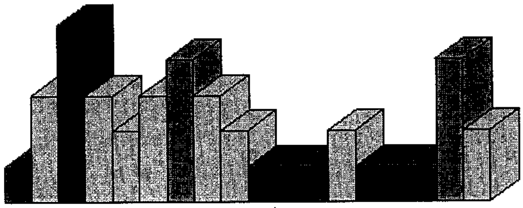


Figure 4 – Probability Distribution

With the real distribution represented with reasonable fidelity in the model, reliance on Gaussian and other analytically manageable distributions can be eased. The machinery in the network to support the distribution of the variable **Intensity** looks like

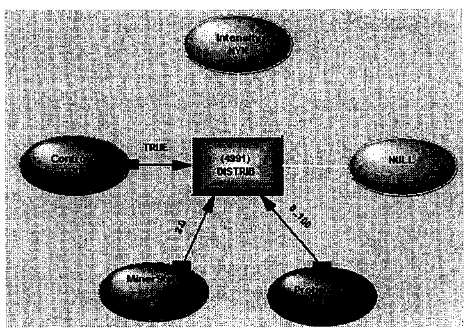


Figure 4 - DISTRIB Operator

Depending on a control state, the DISTRIB operator either "learns" from values arriving at its variable (storing occurrences in different ranges), or makes available a distribution from the values stored within it – the operator responds to logical states being communicated to it and its connections allow information to flow in any direction. The range of the distribution can be controlled by constraints acting on its variable – if X has a range of 1..50 and a distribution on that range, then introducing a constraint such as X < 20 will cut the range and truncate the distribution (it will temporarily put occurrences outside the constrained range at zero).

Where there is complex interrelated information, separate distributions alone do not represent the information adequately. A RELATION operator is used to connect the variables and control the distributions. Figure 5 is a simple example of a two dimensional relation (a maximum of ten dimensions is permissible).

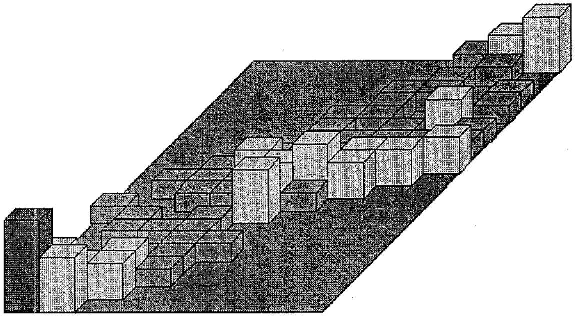


Figure 5 – Building Type vs Earthquake Damage - see [1]

The detailed map that the Relation provides between occurrences in distributions in different dimensions allows the detection of correlations that are smeared away by less detailed representations. The machinery to support the RELATION operator looks like

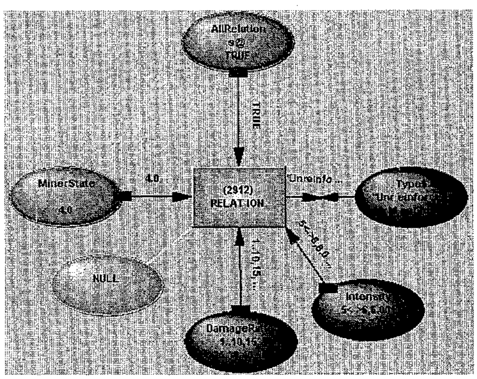


Figure 6 - Relation Operator

A change of distribution at one variable causes a change in distribution at other related variables. In the extreme case, if one variable is set to a singular value, then the other distributions are created by combining the values in the other dimensions corresponding to that value, as Figure 7 shows.

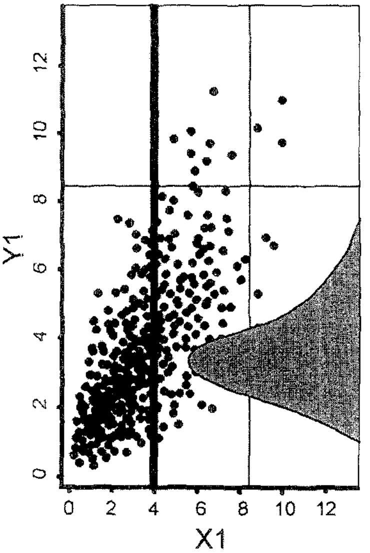


Figure 7 – Singular Value for Relation

Here, setting X1 to 4 has produced a distribution for Y1 based on the available data - real data can be rather sparse, so the contents of the distributions and relations may have been tidied up beforehand. The point is that real data has been used for the transformation between dimensions, rather than analytic approximations. The RELATION is undirected, so asserting a singular value or decreased range for Y1 would create a new distribution for X1. More commonly, the range of one variable would reduce due to some constraint. This would change the distributions of the other variables, which may cause a constraint in another dimension (or group of dimensions) to become active, further reducing distributions related to it. In this way, interdependent risks are handled naturally in closed form analysis, in a manner not dissimilar to the solving of simultaneous equations, except here they are a mixture of simultaneous stochastic and analytic relations.

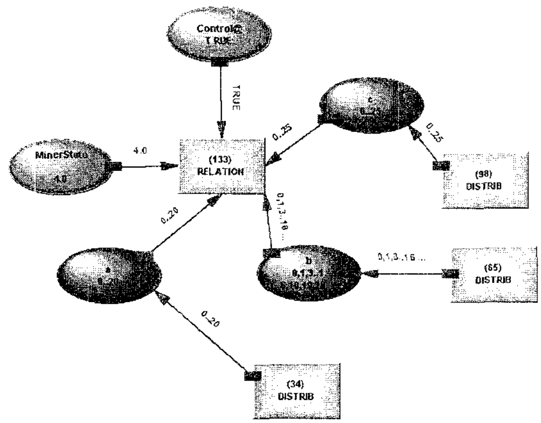


Figure 8 - Stochastic Analog

The structure shown in Figure 8 provides a stochastic analog of the structure shown in Figure 2 for the analytic statement A + B = C. A change in range on any of the variables will affect the distributions of the other variables. The Control connection on the relation mirrors the logical connection on the EQUALS operator, allowing the operator to be turned on and off.

Other operators in the network, triggered by changes of state, can extract any desired statistical measures from the dynamic states of the distribution and relation operators. RANDOM operators, when operating on variables with distributions, will pluck out a value based on the current distribution for that variable and set the variable to that value. Immediately, any variables linked through relation operators will have their distributions reduced, and RANDOM operators acting on them will find a value in their new distributions.

Using relation operators directly between seemingly interdependent risks may not be the most appropriate way to connect them if we have some idea of causality. An example is residential fire claims and motor vehicle claims. If there is asset deflation, we can expect both types to rise due to fraudulent claims. However, if there is a period of abnormally low precipitation, we can expect fire claims to rise because of wildfires and motor vehicle claims to fall because of dry roads. The example also undermines the static view of risk that is usually taken on the liability side – drought may require several years to set the scene for wildfires. If we can find causes for change, and these causes flow to several risks, our results will be much more precise if we manipulate causal variables than if we smear several causes by linking directly between risks. We have already mentioned that distributions and relations have a Learn state, where data is fed into them. There can be an arbitrary amount of analytic structure between the input data and the distributions/relations, allowing causal structure to be hypothesized and validated during the Learn phase. The ability to freely mix analytic and stochastic structure allows for steadily increasing precision, the analytic structure slowly encroaching on the stochastic structure as more is understood. Equally, analytic structure can extend the reach of stochastic structure where there is no history. An example of a complex risk, described by a combination of analytic and stochastic structure and requiring continual update of the structure, is given in [1].

We have shown how stochastic information may be embedded in a model to represent interrelated risks, but the conference call specifically describes the problem of combining stochastic and hypothetical risks. This may occur in several ways:

1. Extension or overlaying of historical risk with hypothetical risk
2. Hypothetical risk only

and the coupling of these two cases in any combination.

Take the example shown, where stochastic information is used inside the rectangle A, a mixture of stochastic and hypothetical is used inside rectangles B and C, and only hypothetical risk is used inside rectangle D, where any historical data may be nonexistent or out of date (in other words, one large disaster has rendered prior experience irrelevant).

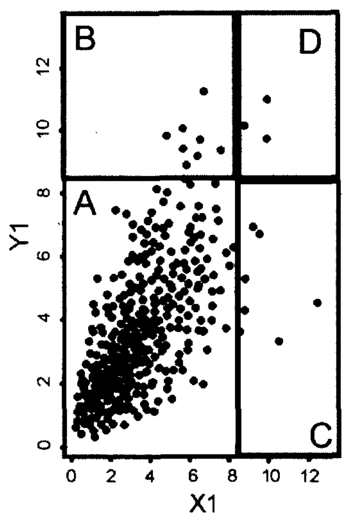


Figure 9 - Hypothetical and Historical Risk

The easiest way to handle this is to augment the historical risk or create new risk distributions by using analytic structure to generate new data points in the model, the distribution and relation operators continuing in the Learn state until sufficient detail has been generated for the hypothetical component of risk. If the hypothetical risk needs to respond to current conditions in the model, then a different, dynamically switching, approach needs to be used. As an illustration, Figure 10 shows stochastic and analytic probability elements being combined dynamically.

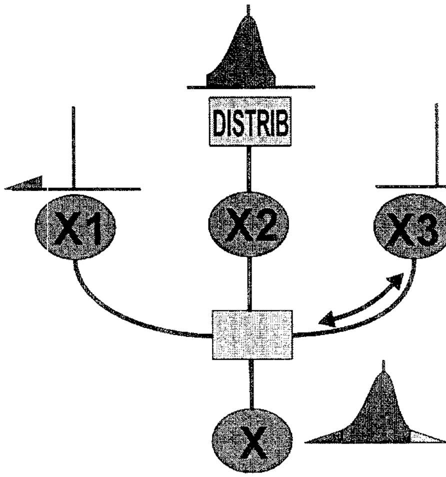


Figure 10 - Dynamically Combining Risk

Activation at any of the components X1 etc. leads to activation of the distribution operator connected to the variable X, and vice versa. This diagram illustrates the undirected nature of the structure, influences flowing at a low level wherever they will. With a little more modeling effort, the crossover points between the stochastic and calculated components of the risk can also be dynamically determined.

A Dynamic Environment needs Knowledge

Actuarial analysis has for a long time relied on the relative stability of the data being analyzed. Models could be built with little concern for maintainability – there certainly would be no major change in their structure over their life. Recent events have demonstrated how rapidly new risks may appear, and even if their primary effects can be avoided by rewriting policies, their secondary effects cannot. It is now desirable to have models which can be quickly adapted to changing circumstances (it always was, we needed a shock to remind us). If we look at the programmatic approach to building models, we see the cognitive scaffolding (the modeler) being used to build the model, and then being removed from it before the model is put into operation. The modeler needs to have anticipated any change in topology in the model and provided instructions to handle these. New risks, or the realization of some interconnectedness among existing risks, will often invalidate the topology the modeler has constructed, resulting in slow adaptation to change (or resistance to change because of the large intellectual investment in the existing model, and the sensitivity of programmatic models to topological change).

An alternative approach is to use the undirected knowledge network structures we have described. Each element in the structure determines direction of flow dynamically, so changes in topology can be made without requiring overall dataflow to be recast in the modeler's head. The interconnecting logic is basically sentential (extended to handle errors and unknowables), allowing the model to reason about what is happening (and reducing the distance between our understanding of the model's operation and our thinking about what needs to be done). That is, not so much of the cognitive support structure is removed when the model is placed in operation. The undirectedness of the structure results in the property of extensibility – structures can be combined easily because the phasing of the operation is implicit in the elements of the component structures.

As we have shown, embedding activity in the structure makes it simple to combine analytic and stochastic knowledge, and allow them to interoperate without the crudity of curve-fitting. In reasonably complex applications such as DFA, the analysis times for knowledge structures are comparable with programmatic methods – flow direction in the network is determined dynamically, but will not change unless there is a change to the input of the analysis that determined it.

Learning from Simulation

doing 100,000 simulations ... but I don't think it actually helps you answer some of the fundamental questions. [4]

Simulation, correctly done, should help you answer some of the fundamental questions. If we take the example of a pilot on a flight simulator, it is not desired to have a report which shows that the pilot crashes 10% of the time – instead, it is desired that the pilot change his/her behavior in real situations as a result of simulated exposure to difficult conditions. Similarly in insurance, simulation is not about giving graphs to management, it is about learning from the experience. It is also not about trying to precisely quantify some losses in ten years time, it is working out what management response might be compared to what it should be as certain patterns begin to appear in the market.

If people resolutely refuse to look at what the DFA simulations are telling them (and with good reason - many of the simulations may be nonsensical due to the crudity of the embedded strategy), then perhaps it is appropriate to introduce some machine learning into the simulation. Machine learning may sound esoteric, but it is easily done by using the same distribution/relation operators as hold the stochastic information for the simulations in the first place, and allowing the results of simulation runs to modify their contents, and their contents to be used to control the simulation runs, so the system "learns" what is required to increase profits and avoid ruin. The obvious difficulty with this approach is that management sees from the simulations that ruin is unlikely, without understanding what the simulation model is doing to avoid it.

A reaction to meaningless simulation has resulted in increased interest in closed form analysis such as RAROC. The conceptual difficulty with a one period analysis like RAROC is that the insurance company followed a particular trajectory to reach the start of the period, and that trajectory is embedded in the mental models of management, thus controlling their strategic viewpoint – it is a simulation over a number of periods, with only the last period evaluated explicitly, the prior periods being implicit in the strategy. A company that had reached the same point by following a different trajectory would probably make very different decisions for the same future period. The strategy that is initially embedded in the simulation should represent current thinking (conditioned by its trajectory), and the simulation can then show how that strategy would operate as the simulation moves from now into the future (and the strategy changes in response to a changing situation). The actual dates into the future are irrelevant, except to act as a brake on how quickly adverse conditions can materialize (and how quickly the mental models of management will change in response to the effects of those adverse conditions).

Conclusion

Knowledge structures can remove many of the obstacles to representation of complex risks. Their undirected nature and active consistency maintenance allows for rapid, controlled, changes to a complex model in the face of changing circumstances. Their intrinsic property of self-modification introduces a dynamic structure, an ability to represent complex strategy and a self-learning ability to DFA for the first time.

References

[1] Various (2002) *AcKnowNet Active Knowledge Manager – Final Report (abridged), (IST-2001-32533)*, retrieved March 2003 from http://www.inteng.com.au/Defence/abridged_report.htm

[2] Brander, J and Dawe, M *Whither Directionality in Machine Based Analysis*, retrieved March 2003 from http://www.inteng.com.au/whither.htm

[3] Brander, J, *A Machine is a Machines*, retrieved March 2003 from http://www.inteng.com.au/Defence/machine.htm

[4] Brehm, P., *RAROC Case Study,* ERisk iConference, June 2001

[5] Embrechts P., McNeil A.J. and Straumann D.: *Correlation and dependence in risk management: properties and pitfalls.* In Risk management: value at risk and beyond, edited by Dempster M. and Moffatt H.K., Cambridge University Press (2002)

The Aggregation and Correlation of Insurance Exposure

Glenn G. Meyers, FCAS, MAAA
Fredrick L. Klinker, FCAS, MAAA, and
David A. Lalonde, FCAS, MAAA, FCIA

The Aggregation and Correlation of Insurance Exposure

By

Glenn Meyers – Insurance Services Office

Fredrick Klinker – Insurance Services Office

David Lalonde – AIR Worldwide Corporation

Abstract

This paper begins with a description of how to calculate the aggregate loss distribution for an insurer. The model underlying this calculation reflects dependencies between the various lines of insurance. We include most of the standard insurance exposures as well as property catastrophe exposure. Next we show how this aggregate loss distribution can be used to allocate the cost of capital and evaluate various reinsurance strategies. We demonstrate the use of this methodology on two illustrative insurers. We believe this methodology can be used in practice by most insurers.

1. Introduction

This paper has three objectives:

1. Demonstrate a practical method to determine the distribution of an insurer's aggregate loss payments. This includes not only losses from the contracts it currently is insuring, but also contracts that have expired but still have outstanding claims. This distribution will depend on the variation of each contract's claim frequency and severity. It will also reflect dependencies among the various hazards insured.
2. Using the results of Objective #1, demonstrate how to determine the amount of capital needed for an insurance company based on its risk of loss.
3. Using the results of Objective #2 demonstrate how to allocate the cost of capital to lines of insurance and evaluate given reinsurance strategies.

We will illustrate the use of our model and methodology on two illustrative insurance companies. The parameters for the loss models were obtained from analyses by Insurance Services Office (ISO) and AIR Worldwide Corporation (AIR).

We treat the time value of money by assuming a fixed risk-free interest rate. While the assets of an insurer are not always risk-free, a full treatment of asset risk is beyond the scope of this paper. Thus, we should expect insurers to have more capital than that indicated by the methodology described in this paper because they have asset risk.

We begin with a description of possible ways to model an insurer's distribution of underwriting losses. This description will include ways to model dependencies among the various insurance lines of insurance. It will also discuss how to parameterize these models.

Next we will describe how we calculate the required capital. This description will include a short survey of the issues involved in making such a calculation. It turns out that there is no strong consensus on how to do this; but, if we are to get a final answer, we must and do pick one method.

We then move on to developing a methodology for allocating the cost of capital to each line of insurance. As we do in our section on calculating the required capital, we will include a short survey of the issues involved in doing this. Again we note that there is no strong consensus on how to do this but, as before, we do pick one method.

While we recognize that others may differ in their methodology for solving these problems, we do feel that our methodology for calculating both the required capital and allocating the cost of capital to lines of insurance is reasonable. We note that the underwriting risk model that we have built to solve these problems could be used for other methodologies.

2. Models of Insurer Losses

This section begins with a description of the classic collective risk model, and it then enhances it with correlations or, more precisely, dependencies generated by parameter uncertainty.

Next we introduce catastrophe models, in which the dependencies are caused by geographic proximity. We describe catastrophes generated by hurricanes and earthquakes.

2.1 The Collective Risk Model

The collective risk model (CRM) describes the total insured loss in terms of the underlying claim count and claim severity distributions for each line of insurance. We describe this model by the following simulation algorithm.

Simulation Algorithm #1

Step

1. For each line of insurance, h, with uncertain claim payments, do the following:
   - Select random claim count K_h from a distribution with mean λ_h, where λ_h is the expected claim count for line of insurance h.
   - For each h, select random claim sizes, Z_{hk}, for $k = 1,\ldots, K_h$.
2. Set $X_h = \sum_{k=1}^{K_h} Z_{hk}$ = Loss for line of insurance h.
3. Set $X = \sum_h X_h$ = Loss for the insurer.

This formulation of the CRM assumes that the losses for each class are independent. We now introduce a dependency structure into the CRM with the following algorithm.

Simulation Algorithm #2

Step

1. For each line of insurance h with uncertain claim payments, do the following:
   - Select a random claim count K_h from a distribution with mean λ_h, where λ_h is the expected claim count for line of insurance h.
   - For each h select a random claim size, Z_{hk}, for $k = 1,\ldots,K_h$.
2. Set $X_h = \sum_{k=1}^{K_h} Z_{hk}$ = Loss for line of insurance h.
3. Select a random β from a distribution with $E[\beta] = 1$ and $Var[\beta] = b$.
4. Set $X = \beta \cdot \sum_h X_h$ = Loss for the insurer.

The extra step of multiplying all the losses by a random β adds variability in a way that losses for each line of insurance will tend to be higher, or lower, together at the same time. This induces one kind of dependency, or correlation, among the losses of different lines of insurance. One can think of b as a parameter that quantifies the uncertainty in the economic environment affecting multiple lines of insurance.

Figures 1-4 provide a graphic illustration of how Simulation Algorithm #2 generates dependency and correlation. In these figures, we randomly selected X_1 and X_2. Next we randomly selected β. We then plotted βX_1 against βX_2. If we do not change the distributions X_1 and X_2, a higher b will lead to a higher coefficient of correlation. But, as illustrated in Figures 3 and 4, the coefficient of correlation also depends on coefficients of variation (CV) of X_1 and X_2.

Figure 1

X_1 and X_2 are independently drawn random variables with CV=0.1.

β was drawn from a distribution with b=Var[β] = 0. Thus ρ = 0.00.

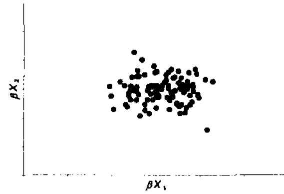


Figure 2

X_1 and X_2 are independently drawn random variables with CV=0.1

β was drawn from a distribution with b=Var[β] = 0.005. Thus ρ = 0.33.

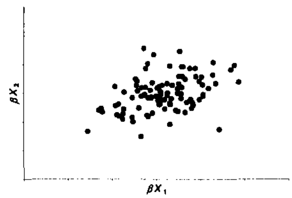


Figure 3

X_1 and X_2 are independently drawn random variables with CV=0.1

β was drawn from a distribution with b=Var[β] = 0.020. Thus ρ = 0.66.

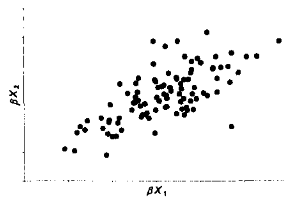
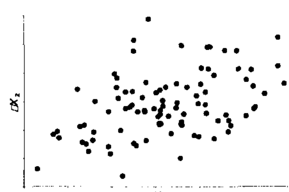


Figure 4

X_1 and X_2 are independently drawn random variables with CV=0.2

β was drawn from a distribution with b=Var[β] = 0.020. Thus ρ = 0.33.



Having described one method to introduce dependencies into the collective risk model, we now apply this method to a model of the underwriting losses for an insurer. Here is a summary of the main features of this model.

- It is necessary to hold capital for uncertain losses in expired insurance contracts. Thus the model treats unpaid losses from both new and expired insurance contracts from prior accident years
- We use separate parameter uncertainty multipliers for both claim frequency and claim severity. For line of insurance h, a random claim frequency multiplier, α_h, is applied to the expected claim count parameter, λ_h. Each α_h has a mean of one and a variance of g_h. We call g_h the covariance generator for line of insurance h.
- Each line of insurance is assigned to a distinct "covariance group" according to the line of insurance that it covers. Within a given covariance group, the random claim frequency multipliers, α_h, are identical within line of business, not necessarily identical to other lines of insurance in the same covariance group, but they increase and decrease together.
- The random claim severity multiplier, β, is applied uniformly across lines of insurance.
- One can informally classify the sources of risk in this model into ***process risk*** and ***parameter risk***. Process risk is the risk attributable to random claim counts and claim sizes, and parameter risk is the risk attributable to the randomness of the claim frequency multipliers and the claim severity multiplier.
- When parameter risk operates on several lines of insurance simultaneously, we say that there is correlation generated by parameter risk.

These features are described in the following algorithm.

Simulation Algorithm #3

Step

1. Select a random β from a distribution with $E[\beta] = 1$ and $Var[\beta] = b$.
2. For each covariance group i, select random percentile p_i.
3. For each covariance group i, line of insurance h in the covariance group (denoted by G_i), and accident year y with uncertain claim payments, do the following:
   - Select $\alpha_{hy} = p_i^{th}$ percentile of a distribution with $E[\alpha_{hy}] = 1$ and $Var[\alpha_{hy}] = g_{hy}$
   - Select random claim count K_{hy} from a distribution with mean $\alpha_{hy} \cdot \lambda_{hy}$, where λ_{hy} is the expected claim count for line of insurance h and accident year y in covariance group i.
   1. For each h and y, select random claim size Z_{hyk} for $k = 1, \ldots, K_{hy}$.
4. Set $X_i = \sum_{h \in G_i} \sum_{y} \sum_{k=1}^{K_{hy}} Z_{hyk}$ = Loss for covariance group i.
5. Set $X = \beta \cdot \sum_i X_i$ = Total loss for the insurer.

We now describe our parameterization of this model.

- For the non-catastrophe lines of insurance, we use claim severity distributions derived by ISO. We use a piecewise linear approximation to the ISO models.

- Smaller claims tend to settle quickly. In fitting the models for the distribution of future payments for expired insurance contracts, we removed those claims that are already settled.

- We use the negative binomial distribution to model claim counts. The expected claim count will depend on the insurer's limits and exposure. A second parameter of the negative binomial distribution, called the contagion parameter must be provided. We use estimates of the contagion parameters obtained in an analysis performed by ISO. This analysis is described in the appendix.

- The same analysis in the appendix also provides estimates of the covariance generators, g_h. A noteworthy feature is that these estimates use data from several insurers. This estimation necessarily assumes that each g_h is the same for all insurers writing that particular line of insurance. While we agree in principle that each g_h could differ by insurer, it is unlikely that any single insurer will have enough observations to get reliable estimates of the g_h's.

- The main idea behind the estimation of the parameters, described in the appendix, is that expected values of various statistics that we can calculate from the data are functions of the negative binomial parameters and the covariance generators. We calculated these statistics for a large number of insurance companies and we found parameter values that best fit the statistics we calculated. As we show in the appendix, reliable estimates of these parameters cannot be obtained with data from a single insurer. It is only by combining the data of several insurers that we can obtain reliable estimates of these parameters.

Finally, we describe how we calculate an insurer's distribution of underwriting losses. Since we describe the loss model in terms of a computer simulation, one could actually do the simulations. In practice, many do. We calculate the distribution of underwriting losses with Fourier transforms using the method described by Heckman and Meyers [1983]. The extension of this method to address dependencies is described by Meyers [1999a and 1999b].

Both simulation and Fourier transforms are valid ways to calculate the distribution of underwriting losses. The advantage of Fourier transforms is that one can calculate the distribution of underwriting losses in seconds, where a simulation could take minutes or even hours to do the same task. A disadvantage of Fourier transforms is that it can take a long time to do the initial set-up whereas the set-up time for a simulation is relatively short.

2.2 Catastrophic Perils

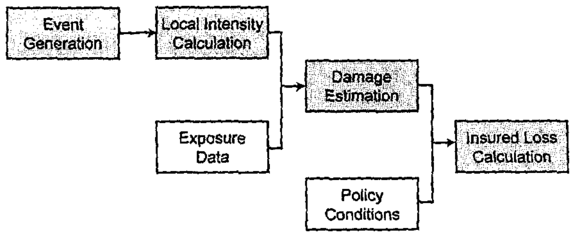
Natural catastrophes such as earthquakes, hurricanes, tornadoes, and floods have an impact on many insureds; and the accumulation of losses to an insurer can jeopardize the financial well-being of an otherwise stable, profitable company. Hurricane Andrew, in addition to causing more than $16 billion in insured damage, left at least 11 companies insolvent in 1992. The 1994 Northridge earthquake caused more than $12 billion in insured damage in less than 60 seconds.

Fortunately, such events are infrequent. But it is exactly their infrequency that makes the estimation of losses from future catastrophes so difficult. The scarcity of historical loss data makes standard actuarial techniques of loss estimation inappropriate for quantifying catastrophe losses. Furthermore, the usefulness of the loss data that does exist is limited because of the constantly changing landscape of insured properties. Property values change, building codes change over time, along with the costs of repair and replacement. Building materials and designs change, and new structures may be more or less vulnerable to catastrophic events than were the old ones. New properties continue to be built in areas of high hazard. Therefore, the limited loss information that is available is not sufficient for directly estimating future losses.

The modeling of catastrophes is based on sophisticated stochastic simulation procedures and powerful computer models of how natural catastrophes behave and act upon the man-made environment. The modeling is broken into four components. The first two components, event generation and local intensity calculation, define the hazard. The interaction of the local intensity of an event with specific exposures is developed through engineering-based vulnerability functions in the damage estimation component. In the final component, insured loss calculation, policy conditions are applied to generate the insured loss.

Figure 5 below illustrates the component parts of the AIR state-of-the-art catastrophe models. It is important to recognize that each component, or module, represents both the analytical work of the research scientists and engineers who are responsible for its design and the complex computer programs that run the simulations.

Figure 5: Catastrophe Model Components (in gray)



2.2a Event Generation Module

The event generation module determines the frequency, magnitude, and other characteristics of potential catastrophe events by geographic location. This requires, among other things, a thorough analysis of the characteristics of historical events.

After rigorous data analysis, researchers develop probability distributions for each of the variables, testing them for goodness-of-fit and robustness. The selection and subsequent refinement of these distributions are based not only on the expert application of statistical techniques, but also on well-established scientific principles and an understanding of how catastrophic events behave.

These probability distributions are then used to produce a large catalog of simulated events. By sampling from these distributions, the model generates simulated "years" of event activity. Many thousands of these scenario years are generated to produce the complete and stable range of potential annual experience of catastrophe event activity and to ensure full coverage of extreme (or "tail") events, as well as full spatial coverage.

2.2.b Local Intensity Module

Once the model probabilistically generates the characteristics of a simulated event, it propagates the event across the affected area. For each location within the affected area, local intensity is estimated. This requires, among other things, a thorough knowledge of the geological and/or topographical features of a region and an understanding of how these features are likely to react to the impact of a catastrophic event. The intensity experienced at each site is a function of the magnitude of the event, distance from the source of the event, and a variety of local conditions. Researchers base their calculations of local intensity on empirical observation as well as on theoretical relationships between the variables.

2.2.c Damage Module

Scientists and engineers have developed mathematical functions called damageability relationships, which describe the interaction between buildings (both their structural and nonstructural components, as well as their contents) and the local intensity to which they are exposed. Damageability functions have also been developed for estimating time element losses. These functions relate the mean damage level as well as the variability of damage to the measure of intensity at each location. Because different structural types will experience different degrees of damage, the damageability relationships vary according to construction materials and occupancy. The model estimates a complete distribution around the mean level of damage for each local intensity and each structural type and, from there, constructs an entire family of probability distributions. Losses are calculated by applying the appropriate damage function to the replacement value of the insured property.

The AIR damageability relationships incorporate the results of well-documented engineering studies, tests, and structural calculations. They also reflect the relative effectiveness and enforcement of local building codes. Engineers refine and validate these functions through the use of post-disaster field survey data and through an exhaustive analysis of detailed loss data from actual events.

2.2.d Insured Loss Module

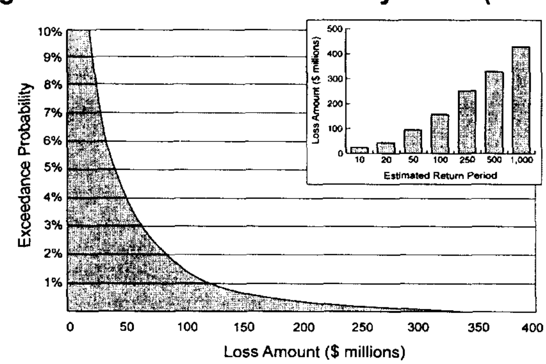
In this last component of the catastrophe model, insured losses are calculated by applying the policy conditions to the total damage estimates. Policy conditions may include deductibles by coverage, site-specific or blanket deductibles, coverage limits and sublimits, loss triggers, coinsurance, attachment points and limits for single or multiple location policies, and risk-specific insurance terms.

2.2.e Model Output

After all of the insured loss estimations have been completed, they can be analyzed in ways of interest to risk management professionals. For example, the model produces complete probability distributions of losses, also known as exceedance probability curves (see Figure 6). Output includes probability distributions of gross and net losses for both annual aggregate and annual occurrence losses. The probabilities can also be expressed as return periods. That is, the loss associated with a return period of 10 years is likely to be exceeded only 10 percent of the time or, on average, in one year out of ten. For example, the model may indicate that, for a given regional book of business, $70 million or more in insured losses would be expected to result once in 50 years, on average, in a defined geographical area, and that losses of $175 million or more would be expected, on average, once every 250 years.

Output may be customized to any desired degree of geographical resolution down to location level, as well as by line of insurance and, within line of insurance, by construction class, coverage, etc. The model also provides summary reports of exposures, comparisons of exposures and losses by geographical area, and detailed information on potential large losses caused by extreme "tail" events.

Figure 6: Exceedance Probability Curve (Occurrence)



2.2.f Correlation

An advantage of this modeling approach is the generation of a stochastic event set that can be used to analyze multiple exposure sets. In this study, individual companies' exposures were analyzed using a common catalog of events. As mentioned earlier, details of insurance programs were also applied, resulting in both net and gross distributions of potential catastrophe losses. By analyzing various sets of exposure against the same set of events we are able to ascertain correlation among the exposure sets.

3. Calculating the Required Capital

This paper is focused on the underwriting risk generated by uncertain loss payments. We assume that all assets are invested at a risk-free rate of return and thus make the simplifying assumption that the capital required by an insurer depends solely on its aggregate loss distribution.

Let X be the random variable for the insurer's total loss. Denote by $\rho(X)$ the total assets that the insurer needs to support its business[1]. Now some of the insurer's assets come from the premium it charges for its business. At a minimum, this amount should equal the expected value of X, $E[X]$. The remaining assets, which we call (economic) capital, must come from investors. We define the capital needed by the insurer by the equation:

$$\text{Capital} = \rho(X) - E[X] \qquad (1)$$

Let α be a selected percentile of X. The tail value-at-risk for X, $TVAR_\alpha(X)$, is defined to be the average of all losses greater than or equal to the α^{th} percentile of X. In this paper we use $\rho(X) = TVAR_{99\%}(X)$.

[1] If we were to allow assets, denoted by A, to be random, we would require A to satisfy $\rho(X-A) = 0$. With translation invariance, this says that $\rho(X) = A$ when A is fixed.

The tail value-at-risk is a member of an important class of risk measures, called coherent measures of risk. These measures are defined by the following set of axioms.

1. Subadditivity — For all random losses X and Y,

$$\rho(X+Y) \leq \rho(X) + \rho(Y).$$

2. Monotonicity — For all random losses X and Y, if $X \leq Y$ for all scenarios, then

$$\rho(X) \leq \rho(Y).$$

3. Positive Homogeneity — For all λ 0 and random losses X,

$$\rho(\lambda X) = \lambda \rho(X).$$

4. Translation Invariance — For all random losses X and constant loss amounts α,

$$\rho(X+\alpha) = \rho(X) + \alpha.$$

These measures were originated by Artzner, *et al.* [1999]. See Meyers [2002] for an elementary description of these measures as well as for other coherent measures of risk.

An insurer can reduce the amount of capital it needs by buying reinsurance. When buying reinsurance, the insurer faces a transaction cost (that is, the reinsurance premium less the provision for expected loss) that replaces a portion of the capital. Note that the insurer does not need to know the reinsurer's pricing assumptions. The insurer can, and perhaps should, use its own estimate of the reinsurer's expected loss to back out the reinsurance transaction cost.

Taxes play an important role in the transaction costs of reinsurance. The insurer deducts reinsurance costs from its taxable income. Capital, whether raised externally or from retained earnings, is subject to corporate income tax. Vaughan [1999] points out that the tendency for reinsurance to stabilize insurer income also provides tax advantages. That gives reinsurance an advantage as a provider of insurer financing.

4. Allocating the Cost of Capital

As noted in the last section, an insurer needs to get capital from investors in order to attract business. The investors want to be compensated for providing this capital at an expected rate of return that is somewhat higher than they would obtain for not exposing their capital to insurance risk. This additional return, or cost of capital, must come from the sum of the premiums charged by each line of insurance. The portion of the cost of capital for an individual insurance contract is often called a risk charge.

Operationally, there are a number of strategies an insurer may take to recover its cost of capital. We list two.

1. Using actuarial formulas, allocate the cost of capital to individual insurance contracts.
2. Allocate the cost of capital to the various lines of insurance and give the line managers the responsibility to recover the cost of capital allocated to their line by whatever combination of pricing and underwriting expertise they can muster.

In this paper we will illustrate the second operating strategy, noting that the second does not necessarily preclude the first. We also note that insurers can purchase reinsurance to reduce their need for capital. In what follows, we will also address the use of reinsurance as part of an underwriting strategy.

Our operating strategy is to establish a target return on the marginal capital for each line of insurance. If the market will not allow the insurer to obtain this target return in a given line of insurance, the insurer should consider tightening its underwriting standards and reducing its exposure in this line of insurance.

We now give our rationale for using this capital allocation formula.

We take it as a given that a sound method of allocating capital should lead to decisions that benefit the entire operation of an insurer.

This discussion will be somewhat informal. A more rigorous treatment of this subject is provided by Meyers [2003]. We shall quote a number of results that are proved in that paper.

Proposition 1

Including a line of insurance in an insurer's portfolio will increase the insurer's expected return on capital if and only if the line of insurance's expected return on marginal capital is greater than the insurer's current expected return on capital.

Proposition 2

Let the insurer's capital be determined by Equation (1), with $\rho(X)$ being a subadditive measure of risk. Then the sum of the marginal capitals for each line of insurance is less than or equal to the insurer's total capital.

As we shall see in the examples below, we expect strict inequality to be common. When this is the case, at least some of the lines of insurance will have an expected return on marginal capital that is greater than the insurer's overall return on capital. However there are conditions when we can prove that the sum of the marginal capitals will be equal to the total capital.

Definition 1

Suppose for line of insurance h, the random losses, X_h, are equal to a random number, U_h, times the exposure measure, e_h, for all possible values of e_h. Then, following Mildenhall [2002], the distribution of X_h is said to be ***homogeneous*** with respect to the exposure measure, e_h.

Proposition 3

Assume that the needed capital is a smooth (differentiable) function of the exposure.

Let the random loss, X_h, for the h^{th} line of insurance be a homogeneous random variable for each contract with respect to some exposure measure, e_h. Let $X = \sum_h X_h$. Let the insurer's capital be determined by Equation (1), with $\rho(X)$ being a measure of risk satisfying the positive homogeneity axiom. Then the sum of the marginal capitals for each line of insurance is equal to the insurer's total capital.

An early version of Proposition 3, assuming each X_i has a lognormal distribution and using a different formula for calculating the needed capital, was proved by Myers and Read [2001]. Mildenhall [2002] proved that the homogeneity assumption was both necessary and sufficient for the Myers-Read result. The proof of Proposition 3 above is a direct consequence of Lemma 2 in Mildenhall's paper.

Note that the definition of homogeneity bears a strong resemblance to the way we introduce parameter risk in Section 2 above. As the exposure (in Section 2, quantified by the expected claim counts λ_{hy}) increases, the parameter risk becomes an increasingly large part of the total risk. But in the parameterization of our model, the parameter risk is rarely dominant enough to assume homogeneity.

Proposition 4

Assume that the needed capital is a smooth (differentiable) function of the exposure. If we can continuously adjust the exposures while holding the needed capital constant, the maximum expected return on capital occurs when the expected return on marginal capital is the same for all lines of insurance.

Note that Proposition 4 does not require homogeneity with respect to some measure of exposure. If the loss random variables are not homogeneous, the equal expected returns on marginal capital under the optimality conditions of the proposition will be higher than the insurer's overall return on capital.

Definition 2

The ***heterogeneity multiplier, HM***, for an insurer is its needed capital divided by the sum of the marginal capitals for each line of insurance.

The motivation for this definition arises from the fact that most insurers will have a total capital that is higher than the sum of the marginal capitals for each line of insurance. In theory, a market could evolve with very large insurers where parameter risk dominates the process risk, and the homogeneity conditions required by Proposition 3 would be reasonable. In practice, the distribution of losses for lines of insurance are far from homogeneous, and the heterogeneity multiplier for a given insurer will be noticeably higher than the theoretical minimum of 1.

Our allocated capital will be equal to the marginal capital times the insurer's heterogeneity multiplier. To summarize, the rationale for this is based on:

1. Proposition 4 – The expected return on marginal capital should be equal for all lines of insurance if the insurer is to make the most efficient use of its capital.
2. Propositions 2 and 3 – The sum of the marginal capitals over all lines of insurance is less than or equal to the total capital. The conditions that will force equality are not satisfied.

Note that the rationale for our allocation formula charge depends on the individual insurance contracts being a small part of an insurer's portfolio, so that the smoothness criterion of Proposition 3 and 4 is a reasonable assumption.

Finally, note that the insurer's pledge to pay losses can be a long-term commitment. As time goes on, the insurer pays some losses and the uncertainty in future loss payments declines. Therefore the insurer can release some of the original capital allocated to a line of insurance.

In the current year, the insurer will have its capital supporting the outstanding losses from prior accident years. We can apply the logic described above and allocate capital to outstanding loss reserves. We calculate the reduction in needed capital when the outstanding losses are removed from the insurance company, and then allocate the capital in proportion to the marginal capital of each underwriting division and each loss reserve. Keep in mind that when establishing target rates of return for the current year, we must consider how much capital the insurer will allocate to the outstanding losses in future years. To do that, the insurer needs a plan for its future business.

5. The Cost of Financing Insurance

As noted above, an insurer must be able to pay its cost of capital out of the premiums charged to the insureds. Now the cost of capital is also affected by reinsurance and the returns on invested assets. Informally, we call the net cost of capital, reinsurance and investment income, ***the cost of financing insurance***. In this section, we show how to draw upon the considerations listed above to calculate this cost.

Investors provide the capital to the insurer. In return, they expect to receive a cash flow reflecting:

1. Premium income,
2. Payments to reinsurers,
3. Investment income,
4. Loss and expense payments, and
5. Income from the capital that is released as liabilities either expire or become certain.

Premium income and payments to reinsurers contain provisions for losses and expenses. It will simplify matters to remove loss and expense payments from our immediate attention by taking expected values and allowing the actual losses in (4) to cancel out the expected loss provisions in (1) and (2). That simplification allows us to concentrate on the cash flow of insurer capital and the net cost of reinsurance; that is, the cost of financing insurance.

After netting out the insurer's loss and expense payments, the investors receive a cash flow reflecting:

1. Income from the profit provision in the premium,
2. Payments of the net costs to reinsurers,
3. Investment income from the capital held for uncertain liabilities, and
4. Income from the capital that is released as liabilities either expire or become certain.

Based on input from the its board of directors, the insurer establishes a target rate of return, r, on its capital. It makes its targeted return on capital if the present value of that cash flow, evaluated *at the targeted return on capital,* is equal to the invested capital. If we allow that:

1. The insurer collects the profit provision in the premium immediately.
2. The insurer makes its reinsurance payments immediately.
3. The insurer determines its necessary capital at the beginning of the year and holds that capital at the end of the year. The insurer then releases capital not needed for the next year. The insurer simultaneously releases investment income on the invested capital.

Then the profit provision necessary for the insurer to make its targeted return on equity is equal to:

Capital + Reinsurance Transaction Costs – Present Value of Released Capital

To get the profit provision for each underwriting division we need to calculate the marginal cost of capital and the transaction costs for reinsurance for: (1) each underwriting division; and (2) each outstanding loss reserve. We now examine the calculations in some detail in the table below.

Table 1

Component for Accident Year *y*	Symbol
Capital investment for current calendar year $y+t$ Note: The insurer needs the capital to cover claims from the current year as well as claims incurred in prior years. The capital also covers business projected for accident years, up to and including year $y+t$.	$C(t)$
Capital needed in calendar year $y+t$ if the insurer removes line of insurance h and accident year y	$C_h(t)$
Marginal Capital for line of insurance h in calendar year $y+t$	$\Delta C_h(t) \equiv C(t) - C_h(t)$
Sum of marginal capitals in calendar year $y+t$	$SM(t)$
Heterogeneity Multiplier	$HM(t) \equiv C(t)/SM(t)$
Capital allocated to line of insurance h for calendar year $y+t$	$A_h(t) \equiv \Delta C_h(t)\times HM(t)$
Transaction costs for line of insurance h's reinsurance (for current accident year only)	$R_h(0)$
Profit provision for line of insurance h	$\Delta P_h(0)$
Insurer's return on its investments	i
Insurer's target return on capital	r

The capital allocated to a given time period earns interest until the beginning of the next period. At that time, the insurer releases a portion of the capital either to pay for losses or to return to the investors.

Table 2

Time	Financial Support Allocated at Time t	Amount Released at Time t
0	$A_h(0) + R_h(0)$	0
1	$A_h(1)$	$Rel_h(1) = A_h(0)(1+i) - A_h(1)$
---	---	---
t	$A_h(t)$	$Rel_h(t) = A_h(t-1)(1+i) - A_h(t)$
---	---	---

Then:

$$\Delta P_h(0) = A_h(0) + R_h(0) - \sum_{t=1}^{\infty} \frac{Rel_h(t)}{(1+r)^t}. \qquad (2)$$

Equation 2 gives the profit provision; i.e. the cost of financing insurance for line of insurance h.

Rearranging the terms of Equation 2 in increasing order of t yields the following simplification.

$$\Delta P_h(0) = (r-i) \cdot \sum_{t=1}^{\infty} \frac{A_h(t)}{(1+r)^t} + R_h(0) = (r-i) \cdot \sum_{t=1}^{\infty} \frac{HM(t) \cdot \Delta C_h(t)}{(1+r)^t} + R_h(0). \qquad (3)$$

Insurers deduct the cost of reinsurance, including the reinsurer's expenses and profit, from taxable income. Given the expected loss ratio, ELR, for a reinsurance contract, the net cost of the reinsurance is then equal to:

$$R_h(0) = \text{Expected Reinsurance Recovery} \times \left(\frac{1}{ELR} - 1 \right) \times (1 - \text{Tax Rate}). \qquad (4)$$

6. Illustrative Examples

We now illustrate the calculation of the cost of financing insurance with two sample insurance companies. We will examine a number of reinsurance strategies.

The following table gives a summary description of each company.

Table 3

	Expected Losses	
	Insurer #1	Insurer #2
Hurricane	10,000,000	1,000,000
Earthquake	5,000,000	500,000
CMP Property	150,000,000	15,000,000
Homeowners	350,000,000	35,000,000
PP Auto Liability	350,000,000	35,000,000
PP Auto Phys Damage	250,000,000	25,000,000
CMP Liability	100,000,000	10,000,000
Total	1,215,000,000	121,500,000

Insurer #1 is a medium sized insurer writing personal and small business coverages. It has some catastrophe exposure. Insurer #2 is similar to Insurer #1 except it has exactly one tenth of its exposure.

In this illustration, we assume that there is no change in the insurers' business plan.

We calculated aggregate loss distributions for each insurer using the ISO Underwriting Risk Model. This model calculates the aggregate loss distributions using the method of Fourier Inversion as described by Heckman and Meyers [1983] and Meyers [1999b]. The underlying claim severity and claim count distributions were derived from an analysis of data reported to ISO.

The catastrophe loss distributions were derived with AIR's CLASIC2 catastrophe model using exposures reported to ISO. We used a composite model based on the combined exposures of ten insurers and multiplied the loss amounts by the factor that yielded the expected losses we selected for these illustrations.

The aggregate loss distributions not only include the losses from the current accident year, but they also include unpaid losses from prior accident years.

Exhibit 1 gives the aggregate loss distribution for each insurer when no reinsurance was purchased. The exhibit provides the value-at-risk and the tail value-at-risk for a variety of levels. We also calculated the capital implied by Equation 1 at each level. In subsequent calculations of the cost of capital, we will use the 99% level. It is worth noting that while the exposure underlying Insurer #1 is ten times the exposure underlying that of Insurer #2, the capital implied by $TVaR_{99\%}$ is only 5.4 times as much. This is a reflection of the greater diversification obtained by larger insurers.

The next step is to calculate the marginal capital for each line of insurance and accident year. Recall that prior accident years with unsettled claims contribute to the need for capital. The process proceeds by calculating the required capital with each line/accident year removed in turn, and then calculating the marginal cost of capital by subtracting the calculated capital from the original capital. We next calculate the heterogeneity multiplier by dividing the total capital by the sum of the marginal capitals. The allocated capital is then set equal to the marginal capital times the heterogeneity multiplier, with the result that the capital allocated to each line of insurance adds up to the capital required for each insurer.

Next we calculate the cost of capital by line of insurance using Equation 3. In our example we chose the insurer's expected rate of return, $r = 15\%$, and the return on investments, $i = 6\%$. When there is no reinsurance, we can calculate the cost of capital directly from Exhibit 2. Let's work through CMP Property for Insurer #1 as a sample calculation. The capital allocated to CMP Property for the current year is $46,464,160. Since we are assuming no change in the insurer's business plan, next year we expect to allocate $16,306,206 to the loss reserve. By Equation 3, the cost of capital is equal to (15% – 6%)×(46,464,160/1.15+16,306,206/1.15^2) which is equal to $4,746,011.

The remaining cost of capital calculations are on Exhibit 3 for Insurer #1, and on Exhibit 4 for Insurer #2.

Finally, we consider the effect of reinsurance. For Insurer #1 we examine retentions of $50 million and $25 million, with 5% participation in losses above the retention. For Insurer #2 we examine retentions of $5 million and $2.5 million, with 5% participation in catastrophe losses above the retention. We also consider full reinsurance on losses in excess of $1 million on the other lines.

The cost of reinsurance is expressed in terms of the expected loss ratios and is given in Table 4.

Table 4

Reinsurance Expected Loss Ratios

Line of Insurance	Reinsurance Expected Loss Ratio
Hurricane	70%
Earthquake	70%
CMP Property	80%
Homeowners	80%
PP Auto Liability	85%
PP Auto Phys Damage	80%
CMP Liability	90%

The results of the calculations are given in Exhibits 3 and 4. The strategies with the lowest cost of financing insurance are highlighted on the exhibits. It is most efficient for Insurer #1 to purchase catastrophe reinsurance at the $50M retention, and to purchase no reinsurance for the other lines. Insurer #2 should purchase catastrophe insurance at the $5M retention, and also purchase the excess of loss reinsurance for the other lines. In general, it is more advantageous for small insurers to purchase reinsurance.

Exhibit 1

Aggregate Loss Distributions with No Reinsurance

	Insurer #1			Insurer #2	
	Aggregate Mean	2,214,538,724		Aggregate Mean	221,453,872
	Aggregate Std. Dev	150,248,026		Aggregate Std. Dev	22,055,783

Probability Level	Value at Risk	Tail Value at Risk	TVaR Implied Capital	Value at Risk	Tail Value at Risk	TVaR Implied Capital
0.00000	0	2,214,538,724	0	0	221,453,872	0
0.50000	2,207,316,395	2,333,813,560	119,274,837	218,924,759	238,276,926	16,823,054
0.55000	2,226,144,404	2,346,827,330	132,288,606	221,482,225	240,286,095	18,832,222
0.60000	2,245,518,205	2,360,713,596	146,174,872	224,155,556	242,471,170	21,017,298
0.65000	2,265,785,629	2,375,744,474	161,205,751	227,018,434	244,886,204	23,432,332
0.70000	2,287,386,709	2,392,304,906	177,766,183	230,132,839	247,610,427	26,156,555
0.75000	2,311,077,498	2,410,974,588	196,435,865	233,628,277	250,765,826	29,311,954
0.80000	2,338,249,056	2,432,691,783	218,153,060	237,722,551	254,556,639	33,102,766
0.85000	2,370,074,024	2,459,213,540	244,674,816	242,754,714	259,372,335	37,918,463
0.90000	2,410,647,140	2,494,364,537	279,825,813	249,559,535	266,095,212	44,641,340
0.92500	2,438,438,516	2,518,002,646	303,463,922	254,309,896	270,858,293	49,404,421
0.95000	2,474,058,835	2,549,901,446	335,362,723	260,925,559	277,612,418	56,158,545
0.95500	2,481,966,185	2,557,988,890	343,450,166	262,578,399	279,383,801	57,929,928
0.96000	2,491,812,025	2,566,890,844	352,352,121	264,478,659	281,368,941	59,915,068
0.96500	2,502,702,587	2,576,859,477	362,320,753	266,672,502	283,630,336	62,176,464
0.97000	2,515,539,517	2,588,212,358	373,673,634	269,180,787	286,257,956	64,804,084
0.97500	2,529,863,230	2,601,455,650	386,916,927	272,133,993	289,390,878	67,937,006
0.98000	2,546,973,953	2,617,391,407	402,852,684	275,871,191	293,261,157	71,807,284
0.98500	2,568,686,470	2,637,546,022	423,007,298	280,633,668	298,317,523	76,863,651
0.99000	2,597,771,711	2,665,306,927	450,768,203	287,622,598	305,543,931	84,090,059
0.99500	2,650,836,978	2,711,022,014	496,483,291	300,182,048	317,998,993	96,545,121
0.99900	2,757,830,730	2,813,156,828	598,618,105	329,317,133	345,279,002	123,825,129
0.99950	2,795,656,351	2,856,560,534	642,021,811	340,757,188	356,147,056	134,693,184
0.99990	2,896,155,347	2,953,848,239	739,309,515	365,548,241	380,674,424	159,220,551

Exhibit 2

Capital Allocation with No Reinsurance

	Insurer #1			Insurer #2		
Line of Insurance	Expected Loss	Marginal Capital	Allocated Capital	Expected Loss	Marginal Capital	Allocated Capital
Hurricane1	10,000,000	8,036,268	12,092,714	1,000,000	221,927	538,230
Earthquake1	5,000,000	7,821,147	11,769,006	500,000	201,836	489,505
CMP Property1	150,000,000	30,877,971	46,464,160	15,000,000	7,345,395	17,814,470
CMP Property2	62,204,206	10,836,364	16,306,206	6,220,421	2,832,793	6,870,251
Homeowners1	350,000,000	108,304,574	162,973,181	35,000,000	3,524,303	8,547,341
Homeowners2	105,073,478	18,586,085	27,967,733	10,507,348	725,589	1,759,741
Homeowners3	30,656,934	4,067,161	6,120,131	3,065,693	168,620	408,947
Homeowners4	16,983,825	1,913,653	2,879,603	1,698,382	81,755	198,278
Homeowners5	9,863,946	964,018	1,450,623	986,395	42,378	102,778
PP Auto Liability1	350,000,000	23,947,855	36,035,950	35,000,000	1,204,153	2,920,380
PP Auto Liability2	212,745,098	12,944,523	19,478,495	21,274,510	721,967	1,750,957
PP Auto Liability3	105,927,476	5,754,783	8,659,609	10,592,748	344,588	835,714
PP Auto Liability4	56,129,303	2,831,150	4,260,222	5,612,930	175,818	426,404
PP Auto Liability5	28,308,824	1,360,604	2,047,392	2,830,882	89,366	216,737
PP Auto Phys Dam1	250,000,000	15,568,762	23,427,364	25,000,000	653,326	1,584,484
PP Auto Phys Dam2	19,455,630	973,945	1,465,560	1,945,563	42,236	102,433
CMP Liability1	100,000,000	9,852,146	14,825,187	10,000,000	2,995,816	7,265,623
CMP Liability2	92,040,513	8,867,464	13,343,470	9,204,051	2,813,484	6,823,422
CMP Liability3	90,251,701	8,886,460	13,372,054	9,025,170	3,078,438	7,466,002
CMP Liability4	67,617,504	6,652,022	10,009,745	6,761,750	2,574,079	6,242,803
CMP Liability5	47,175,325	4,801,482	7,225,113	4,717,532	2,106,852	5,109,658
CMP Liability6	32,891,244	3,414,440	5,137,938	3,289,124	1,607,077	3,897,575
CMP Liability7	22,213,719	2,300,466	3,461,666	2,221,372	1,121,335	2,719,527
Total	2,214,538,724	299,563,343	450,773,121	221,453,872	34,673,134	84,091,259
Heterogeneity Multiplier		1.5048			2.4253	

Note: The numeric identifiers following the lines of insurance denote losses from different accident years. The identifier "1" denotes the current accident year, the identifier "2" denotes the first prior accident year, and so on.

Exhibit 3

The Cost of Financing Insurance – Insurer #1

	Cost of Capital	Net Cost of Reinsurance	Cost of Financing	Cost of Capital	Net Cost of Reinsurance	Cost of Financing
	No Reinsurance			Cat Re $50M Retention + Other Re		
Total	32,763,664	0	32,763,664	29,211,706	6,342,331	35,554,037
Hurricane	946,386	0	946,386	221,601	419,873	641,474
Earthquake	921,053	0	921,053	132,903	379,400	512,303
CMP Property	4,746,011	0	4,746,011	3,520,987	3,409,584	6,930,571
Homeowners	15,232,964	0	15,232,964	17,052,018	153,875	17,205,893
PP Auto Liability	4,969,052	0	4,969,052	4,656,223	893,442	5,549,666
PP Auto Phys Damage	1,933,182	0	1,933,182	1,905,098	0	1,905,098
CMP Liability	4,015,016	0	4,015,016	1,722,876	1,086,156	2,809,032
	Cat Re – $50M Retention			Cat Re $25M Retention + Other Re		
Total	**31,741,208**	**799,273**	**32,540,481**	29,071,528	7,015,066	36,086,594
Hurricane	**212,832**	**419,873**	**632,706**	105,010	862,220	967,230
Earthquake	**126,361**	**379,400**	**505,761**	56,234	609,789	666,023
CMP Property	**4,782,691**	**0**	**4,782,691**	3,519,627	3,409,584	6,929,211
Homeowners	**15,731,218**	**0**	**15,731,218**	17,125,589	153,875	17,279,464
PAL	**4,953,994**	**0**	**4,953,994**	4,645,715	893,442	5,539,157
PAPHD	**1,926,125**	**0**	**1,926,125**	1,900,753	0	1,900,753
CMP Liability	**4,007,986**	**0**	**4,007,986**	1,718,599	1,086,156	2,804,755
	Cat Re – $25M Retention			Other Reinsurance Only		
Total	31,613,327	1,472,009	33,085,335	30,419,819	5,543,057	35,962,877
Hurricane	102,816	862,220	965,035	1,084,992	0	1,084,992
Earthquake	54,528	609,789	664,316	1,069,829	0	1,069,829
CMP Property	4,788,789	0	4,788,789	3,543,709	3,409,584	6,953,293
Homeowners	15,781,597	0	15,781,597	16,278,011	153,875	16,431,886
PAL	4,951,928	0	4,951,928	4,740,974	893,442	5,634,416
PAPHD	1,925,236	0	1,925,236	1,941,039	0	1,941,039
CMP Liability	4,008,433	0	4,008,433	1,761,266	1,086,156	2,847,422

Exhibit 4

Cost of Financing Insurance – Insurer #2

	Cost of Capital	Net Cost of Reinsurance	Cost of Financing	Cost of Capital	Net Cost of Reinsurance	Cost of Financing
	No Reinsurance			**Cat Re $5M Retention + Other Re**		
Total	5,597,928	0	5,597,928	**3,093,867**	**634,233**	**3,728,100**
Hurricane	42,122	0	42,122	**21,919**	**41,987**	**63,906**
Earthquake	38,309	0	38,309	**13,088**	**37,940**	**51,028**
CMP Property	1,861,717	0	1,861,717	**398,989**	**340,958**	**739,947**
Homeowners	827,680	0	827,680	**1,657,921**	**15,387**	**1,673,309**
PP Auto Liability	428,804	0	428,804	**551,309**	**89,344**	**640,653**
PP Auto Phys Damage	130,974	0	130,974	**192,279**	**0**	**192,279**
CMP Liability	2,268,323	0	2,268,323	**258,363**	**108,616**	**366,979**
	Cat Re $5M Retention			Cat Re $2.5M Retention + Other Re		
Total	5,567,574	79,927	5,647,502	3,080,666	701,507	3,782,173
Hurricane	13,168	41,987	55,155	10,466	86,222	96,688
Earthquake	7,560	37,940	45,500	5,586	60,979	66,565
CMP Property	1,877,296	0	1,877,296	399,301	340,958	740,260
Homeowners	827,843	0	827,843	1,664,076	15,387	1,679,464
PAL	428,357	0	428,357	550,930	89,344	640,274
PAPHD	130,788	0	130,788	192,106	0	192,106
CMP Liability	2,282,563	0	2,282,563	258,199	108,616	366,815
	Cat Re $2.5M Retention			Other Reinsurance Only		
Total	5,562,310	147,201	5,709,511	3,202,490	554,306	3,756,796
Hurricane	6,691	86,222	92,913	101,108	0	101,108
Earthquake	3,458	60,979	64,437	98,952	0	98,952
CMP Property	1,880,484	0	1,880,484	397,533	340,958	738,491
Homeowners	827,990	0	827,990	1,596,831	15,387	1,612,219
PAL	428,257	0	428,257	554,183	89,344	643,527
PAPHD	130,736	0	130,736	193,607	0	193,607
CMP Liability	2,284,694	0	2,284,694	260,276	108,616	368,892

7. The Effect of Correlation on the Cost of Capital

We began this paper with a discussion of correlation. We now revisit the subject by illustrating the effect that correlation has on our final objective, calculating the cost of capital. To this end, we recalculated the aggregate loss distributions for Insurers #1 and #2 by setting the parameters that govern correlation i.e., the "b" and "g" parameters in Simulation Algorithm 3 equal to zero. The results are in Exhibit 5 and can be compared directly with the results in Exhibit 1. Here are some highlights

Table 5

	With Correlation		Without Correlation	
	Standard Deviation	$TVaR_{99\%}$ Capital	Standard Deviation	$TVaR_{99\%}$ Capital
Insurer #1	150,248,026	450,768,203	109,495,017	332,288,089
Insurer #2	22,055,783	84,090,059	19,502,398	79,579,900

Notice that the effect of correlation is proportionally greater for the larger insurer. This reinforces the point made in Section 2 that parameter uncertainty across lines of insurance has a greater effect when the aggregate loss distribution is relatively less volatile.

Correlation also has an effect on capital allocation. See the following calculations that were done for Insurer #1 in the no reinsurance case.

Table 6

	With Correlation		Without Correlation	
	Cost of Capital	% of Total	Cost of Capital	% of Total
Total	32,763,664		24,341,371	
Hurricane	946,386	2.9%	2,220,791	9.1%
Earthquake	921,053	2.8%	2,286,145	9.4%
CMP Property	4,746,011	14.5%	1,864,473	7.7%
Homeowners	15,232,964	46.5%	11,014,915	45.3%
PP Auto Liability	4,969,052	15.2%	2,294,539	9.4%
PP Auto Phys Damage	1,933,182	5.9%	722,513	3.0%
CMP Liability	4,015,016	12.3%	3,937,996	16.2%

Informally speaking, a line of insurance will need more capital if it has losses at the same time as other lines. In the terminology of Simulation Algorithm 3, CMP Property and Homeowners are in the same covariance group, as are the auto coverages. As a result, they need proportionally more capital.

Exhibit 5

Aggregate Loss Distributions with No Reinsurance and No Correlation

	Insurer #1		Insurer #2	
	Aggregate Mean	2,214,538,724	Aggregate Mean	221,453,872
	Aggregate Std. Dev	109,495,017	Aggregate Std. Dev	19,502,398

Probability Level	Value at Risk	Tail Value at Risk	Implied Capital	Value at Risk	Tail Value at Risk	Implied Capital
0.00000	0	2,214,538,724	0	0	221,453,872	0
0.50000	2,209,849,751	2,301,281,062	86,742,339	218,717,052	236,019,123	14,565,251
0.55000	2,223,502,039	2,310,685,540	96,146,817	220,885,375	237,822,769	16,368,897
0.60000	2,237,531,835	2,320,717,665	106,178,941	223,171,073	239,799,816	18,345,943
0.65000	2,252,191,744	2,331,575,801	117,037,077	225,626,537	242,003,701	20,549,829
0.70000	2,267,811,994	2,343,540,542	129,001,818	228,354,056	244,512,123	23,058,250
0.50000	2,284,846,585	2,357,036,573	142,497,849	231,489,221	247,445,549	25,991,677
0.80000	2,304,030,883	2,372,764,514	158,225,790	235,150,766	251,009,254	29,555,382
0.85000	2,327,886,038	2,391,985,903	177,447,179	239,678,405	255,589,804	34,135,932
0.90000	2,357,616,057	2,417,591,532	203,052,809	246,281,467	262,052,040	40,598,167
0.92500	2,376,043,715	2,434,949,584	220,410,861	250,609,853	266,693,943	45,240,070
0.95000	2,401,438,345	2,458,514,932	243,976,208	257,006,225	273,298,133	51,844,260
0.95500	2,408,119,240	2,464,506,538	249,967,815	258,550,660	275,036,610	53,582,737
0.96000	2,415,016,654	2,471,149,110	256,610,386	260,402,469	276,986,558	55,532,686
0.96500	2,423,124,841	2,478,613,585	264,074,861	262,511,898	279,211,741	57,757,869
0.97000	2,432,195,486	2,487,167,040	272,628,316	264,945,701	281,801,454	60,347,582
0.97500	2,442,666,620	2,497,197,640	282,658,916	267,836,917	284,896,995	63,443,123
0.98000	2,455,285,281	2,509,372,666	294,833,942	271,530,701	288,735,067	67,281,195
0.98500	2,471,124,344	2,524,954,344	310,415,621	276,208,536	293,774,481	72,320,608
0.99000	2,493,190,477	2,546,826,813	332,288,089	283,063,246	301,033,772	79,579,900
0.99500	2,535,366,820	2,584,230,512	369,691,788	295,832,890	313,553,223	92,099,351
0.99900	2,619,031,638	2,675,338,872	460,800,149	324,763,855	339,490,154	118,036,282
0.99950	2,661,125,339	2,716,408,761	501,870,038	335,131,564	349,684,442	128,230,570
0.99990	2,761,594,097	2,810,454,586	595,915,863	358,670,975	373,186,001	151,732,128

8. Summary and Conclusions

This paper started with three objectives:

1. Demonstrate a practical method to determine the distribution of an insurer's aggregate loss payments.
2. Using the results of Objective #1, demonstrate how to determine the amount of capital needed for an insurance company based on its risk of loss.
3. Using the results of Objective #2 demonstrate how to allocate the cost of capital to lines of insurance and evaluate given reinsurance strategies.

We demonstrated our methodology for accomplishing these objectives on two illustrative insurers.

We used the ISO Underwriting Risk Model to determine the aggregate loss distribution. We used the claim count and claim severity distributions provided by the model. For hurricane and earthquake losses, we used the AIR catastrophe model with exposures reported to ISO as input.

Dependencies among the various lines of insurance were reflected in the model by quantifications of parameter uncertainty in the standard lines of insurance and by geographic proximity for the catastrophe exposure.

We believe we have demonstrated that this methodology can be implemented for most insurers.

References

1. Philippe Artzner, Freddy Delbaen, Jean-Marc Eber and David Heath, "Coherent Measures of Risk," Math. Finance 9 1999, no. 3, 203-228 http://www.math.ethz.ch/~delbaen/ftp/preprints/CoherentMF.pdf

2. Philip E. Heckman and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity Distributions and Claim Count Distributions," *PCAS LXX*, (1983) http://www.casact.org/pubs/proceed/proceed83/83022.pdf

3. Glenn Meyers, "Estimating Between Line Correlations Generated by Parameter Uncertainty," *CAS Forum*, Summer (1999a) http://www.casact.org/pubs/forum/99sforum/99sf197.pdf

4. Glenn Meyers, "Discussion of Aggregation of Correlated Risk Portfolios by Shaun Wang," *PCAS LXXXVI*, (1999b) http://www.casact.org/pubs/proceed/proceed99/99705.pdf

5. Glenn Meyers, "Setting Capital Requirements with Coherent Measures of Risk – Part 1 and Part 2," *Actuarial Review*, August and November (2002) http://www.casact.org/pubs/actrev/aug02/latest.htm, http://www.casact.org/pubs/actrev/nov02/latest.htm.

6. Glenn Meyers, "The Economics of Capital Allocation," Presented at the Bowles Symposium, April (2003) http://www.casact.org/coneduc/specsem/sp2003/papers/meyers.doc

7. Stephen J. Mildenhall, "A Note on the Myers and Read Capital Allocation Method," submitted for publication, (2002). http://www.mynl.com/pptp/mr2.pdf

8. Stewart C. Myers and James A. Read, "Capital Allocation for Insurance Companies," *Journal of Risk and Insurance*, December (2001) http://www.aib.org/RPP/Myers-Read.pdf

9. Trent R. Vaughan, "Property Liability Risk Management and Securitization," CAS Discussion Paper Program (1999) http://www.casact.org/pubs/dpp/dpp99/99dpp291.pdf

10. Shaun S. Wang, Virginia R. Young, and Harry H. Panjer, "Axiomatic Characterization of Insurance Prices," *Insurance Mathematics and Economics 21* (1997) 173-182.

Appendix: Estimation of Correlation

Certainly one major driver of actuarial interest in correlation is the perception that positive correlation among lines of business, books of business, etc. has the potential to increase required capital. As a consequence of this observation, it seems to us that the program should be as follows:

- Estimate expected losses or loss ratios,
- measure deviations of the actuals from these expectations,
- and estimate correlations among these deviations as the correlations relevant to the required capital issue.

In an effort to parameterize various ISO models, we have carried out this program. For the sake of parsimony (to limit the required number of parameters to a relative few), we have imposed on correlation a model structure as documented in Meyers [1999a and 1999b]. We estimate correlations within company between lines of business and between company both within and between lines of business. These correlations among companies and among lines of business then drive correlations among insurance contracts written on those companies and lines of business.

Our dataset includes a fairly large number of companies, and our models are parsimonious in the sense of assuming that the same correlation model parameter values apply across all companies within a line of business. So our estimates are in effect pooled estimates. Even so, parameter estimates (contagions and covariance generators) were never more than two or three or four times their associated standard errors. Common statistical practice holds that an estimate is not statistically significant (at the approximately 95% level) unless the estimate in absolute value is at least twice its standard error. Had our dataset not included as many companies or had we attempted to estimate separate parameters by company (or at least by class of company), standard errors would have been larger in relation to their estimates. So it is doubtful that we would have found many parameter estimates significant at the 95% level. The large number of companies and the pooling are necessary to achieve significance.

The next section of this Appendix will address some philosophical issues of just precisely what correlation do we wish to measure anyway, and what are some of the adjustments we must make to observe this correlation. The following section will then discuss the correlation model of Meyers [1999a and 1999b] and an introduction to how we estimate the parameters appearing in the resulting formulae. The remaining sections will discuss the technical details of the estimation, with a few representative results presented at the end. We defer to the end of the model discussion a quick summary of the remainder of this Appendix, because even a quick summary of the technical details requires as background the topics we will discuss in the next two sections.

Correlation of What?

Suppose a realistic forecast, taking into account current rates and prices, estimates of trend, perceptions of current market conditions, etc., indicates that next year's losses will be higher than the long-term average. On the day the business is written, the insurance executive therefore already expects losses higher than average and makes some provision for that. Where the requirement for capital comes from, however, is the recognition that losses could emerge even higher than the already higher expected, and potentially higher than expected simultaneously for a number of lines of business, books of business, etc., due to positive correlation among those books. Thinking in this way clearly identifies the fallacy of measuring correlations of deviations about long-term averages, where some of the deviation from-long term average is due to predictable cycles, trends, etc. What matters, at least for correlation studies relevant to required capital, is not predictable deviation from long-term average but correlated, unpredictable deviation from expectations varying predictably from long-term averages.

As an enlightening thought experiment, consider an optimistic insurance company that consistently forecasts losses lower than their true expected value. Considerably more often than not, deviations of actual from forecast will be positive, yielding apparently fairly significant positive correlations among the outcomes, probably more positive correlation than would result if we were to measure deviations about true expected values. This thought experiment warns us that, to some extent, the correlations we measure will be dependent upon the way we estimate expectations from which we measure deviations.

As a further enlightening thought experiment, we ask what algorithm would most likely produce correlation estimates most relevant to the required capital issue. This would be the algorithm that most closely mimics the actual emergence over time of information in the insurance industry. Suppose for a number of companies and lines of business that we had time series of annual ultimate loss results (or results to date developed to ultimate), as well as potential predictor time series, such as losses emerged at each point in time (not developed to ultimate), rate and price indices, trend estimates at various points in time (based only on information up through that time), indicators of market competitiveness at various points in time, etc. As an example, suppose we sit at the end of year 10 and forecast year 11 based only on what the industry would have known at the end of year 10. Then in year 11 we calculate deviations of ultimate losses from these forecasts. Then we roll the time series forward to the end of year 11 and repeat the process, forecasting year 12, etc. Finally estimate correlations among these deviations.

The problems with this algorithm are at least twofold: 1) We probably need time series with duration of at least a couple of decades--at least the first decade to calibrate the time series forecasting model, plus at least another decade of forecasts from the calibrated model, and their attendant deviations and correlations, so that correlation estimates are not driven too much by events in any one year. In fact, it would probably be useful to have at least a couple of decades of forecasts and deviations so that we could potentially test the stability of correlation estimates over time. 2) We would need to reconstruct time series of what the industry knew at past points in time, such as rate and price indices, past estimates of trend, market competitiveness indices, etc. We might not be able to construct such time series at reasonable cost. Also, we might not be able to reconstruct other time series of what the industry knew or could have known at past times with any reasonable accuracy.

In light of these difficulties, we have constructed "forecasts" about which to measure deviations and correlations via an alternative algorithm. By line of business (LOB) and company, we have about a decade's worth of paid losses developed out to the oldest age in our available loss development triangles. We have not constructed time series of other potential predictors of those loss ratios. Instead, separately by LOB, we have developed generalized additive models for these loss ratios with main effects for company and a non-parametric, non-linear smoother term for year. The year effect is a loess smoother (Not a typo. Loess is a form of localized regression.) of local second degree with smoothing window over years sufficiently wide that long-term trends and turning points are captured without responding much to the random ups and downs of individual years. We have chosen a smoother of local second degree rather than first degree to better respond to turning points in the data.

The downside of this algorithm for correlation analysis is that the use of smoothers produces "forecasts" that, at any given point in time, depend on all of past, present, and future with respect to that point in time. Such "forecasts" may perform better than even the best of forecasts that must depend strictly on only the past, especially with respect to turning points and points of inflection. Therefore, some of what is captured in a smoother-based "forecast" (and therefore considered "predictable" with respect to that forecast) would be unpredictable and not captured by forecasts dependent strictly on the past and would instead be captured in the unpredictable deviations about those forecasts. Therefore, deviations about true forecasts dependent only on the past might tend to be somewhat larger and somewhat more correlated than deviations about smoother-based forecasts. As a consequence, our correlation estimates should be regarded as lower bounds.

On the other hand, the performance of our smoother-based forecasts may not be vastly superior to forecasts based only on the past that take advantage of more information than just losses, such as rate, price, trend, market competitiveness, etc. We would therefore not expect our correlation estimates to be vastly understated. Furthermore, we would expect those correlation estimates to be considerably closer to the mark than estimates based on deviations about long-term averages to the extent that in many of the lines we have studied there has been considerable long-term trend over the last decade; and we would argue that much of this long-term trend was indeed predictable, at least on a rolling one-year-ahead forecast basis.

A Correlation Model Based on Parameter Uncertainty

The reader is referred to Meyers [1999a and 1999b] where one of us has developed a model with correlation driven by parameter uncertainty. The essence of this model is captured in Simulation Algorithm #3 in the main text of this paper. Losses are assumed conditionally independent; but correlation is imposed via severity multipliers assumed common across all lines of business and via frequency multipliers assumed common across all losses within a line of business and at least perfectly correlated, if not identical, across all lines within a so-called "covariance group." This model imposes a certain structure on correlations that depend upon parameters that can be estimated.

Although the models published in Meyers [1999a and 1999b] include both severity and frequency multipliers, we have chosen to fit to a version of the model with just frequency multipliers and have estimated the additional contribution to correlation from severity effects not by fitting data but rather by appeal to our understanding of severity-trend uncertainty. All losses across all lines are assumed multiplied by a common severity multiplier. This multiplier is a random variable with expectation 1 and variance b. If we assume our uncertainty regarding severity-trend translates to an uncertainty regarding severity on the order of 3%, then this translates to a b of approximately $(.03)^2$ 0.001. Although we fit to a model form excluding severity-parameter uncertainty, the data we fit probably includes a component of correlation due to severity uncertainty, because we have certainly made no adjustments to the data to remove this particular uncertainty. Therefore, it is likely that the frequency uncertainty parameters of the model have taken up some of the slack and have responded to both frequency and severity uncertainty, at least to the extent that severity uncertainty can be captured by this model form. Then adding on top of frequency parameters, which may already have captured a portion of the severity effect, a b value estimated from first principles has the potential to overstate the total correlation. This is countervailing to the effect discussed in the previous section of this Appendix, which would potentially cause an understatement of correlation.

We note lastly that we have not yet studied correlations across years. But, within year, we note that we have studied across company/across LOB, across company/within LOB, within company/across LOB, and within company/within LOB (this last would be just variance, the usual process variance but augmented for the additional impact of parameter uncertainty).

Let L_{ijk} be the annual aggregate ultimate loss for line of business i, company j, and year k. Similarly for $L_{i'j'k}$. The two companies j and j' could be the same or different, the two lines i and i' the same or different. Assuming no severity parameter uncertainty, so $b = 0$, the covariance between L_{ijk} and $L_{i'j'k}$ is as given in Meyers [1999a]:

$$Cov[L_{ijk}, L_{i'j'k}] = \delta_{ii'}\delta_{jj'}\left\{\left(\frac{\sigma_i^2}{\mu_i} + \mu_i\right)E_{ijk} + (1+g_i)c_i E_{ijk}^2\right\} + \delta_{G_iG_{i'}}\sqrt{g_i g_{i'}}E_{ijk}E_{i'j'k}. \quad (A.1)$$

- $\delta_{ii'}$ is 1 if and only if $i = i'$ (i.e., the first and second LOBs are the same) and 0 otherwise. Likewise for $\delta_{jj'}$. In other words, the first term is nonzero only when first and second LOBs match, first and second companies match, and first and second years match, in other words, only when calculating variances.

- $\delta_{GiGi'}$ is 1 if and only if the first and second lines of business are in the same covariance group, otherwise 0. To get 1, first and second companies don't have to match, nor do first and second lines of business, but first and second lines of business have to be at least in the same covariance group.

- μ_i and σ_i are the mean and standard deviation of the severity distribution associated with LOB i.

- λ_{ijk} is the expected claim count associated with L_{ijk} and c_i is the contagion for LOB i, so the variance of claim count associated with L_{ijk} is $\lambda_{ijk} + c_i\lambda_{ijk}^2$.

- $E_{ijk} = E[L_{ijk}] = \lambda_{ijk}\mu_i$.

- g_i is the covariance generator associated with LOB i. In other words, in this line of business, parameter uncertainty associated with frequency is captured by a common multiplier across all companies within this line of business, the multiplier being a random variable with mean 1 and variance g_i. The formula above reflects one departure from the referenced Meyers [1999a and 1999b] papers. Whereas those papers assumed the same multiplier across all lines of business within covariance group, it is now assumed that across lines of business within covariance group the frequency multipliers could be different, with different covariance generators, but they are still assumed perfectly correlated. This results in replacing some occurrences of g_i in the earlier formulae with the $\sqrt{g_i g_{i'}}$ appearing above.

Recall that, by definition:

$$Cov[L_{ijk}, L_{i'j'k}] = E[(L_{ijk} - E[L_{ijk}])(L_{i'j'k} - E[L_{i'j'k}])].$$

Define the normalized deviation

$$\Delta_{ijk} = \frac{L_{ijk} - E[L_{ijk}]}{E[L_{ijk}]}.$$

Then divide through equation A.1 above by $E_{ijk}E_{i'j'k}$ to find:

$$E[\Delta_{ijk}\Delta_{i'j'k}] = \frac{\delta_{ii'}\delta_{jj'}\left(\frac{\sigma_i^2}{\mu_i} + \mu_i\right)}{E_{ijk}} + \delta_{ii'}\delta_{jj'}(1+g_i)c_i + \delta_{GiGi'}\sqrt{g_i g_{i'}}\,. \tag{A.2}$$

So, if $i = i'$ and $j = j'$, we are looking at a variance. Then that variance is a regression on $1/E$, with regression coefficient depending only on the parameters of the underlying severity distribution and with intercept term equal to $c_i + g_i + c_i g_i$. This term is approximately $c_i + g_i$ because the product $c_i g_i$ can be expected to be much smaller than either c_i or g_i, both of which are expected themselves to be small. If first and second companies are different but first and second lines of business are the same, then the expectation above is g_i, the covariance generator for the single common line of business. Regardless of whether first and second companies are the same or different, if first and second lines of business are different, then the expectation above becomes $\sqrt{g_i g_{i'}}$, the geometric average of the covariance generators of the two lines of business. If the two lines of business are in different covariance groups, then the expectation above is zero.

Suppose we estimate those expectations, and hence the parameters of our correlation model, from (weighted) averages of or regressions on pairwise products of normalized deviations of our underlying data. We will discuss the appropriate weights later. Consider first all pairwise products of normalized deviations where the first and second LOBs are equal to a single selected LOB of interest, with first and second companies different. From equation A.2, we expect an appropriately weighted average (across all companies and years) of these pairwise products to approximate the expectation g_i. We estimate $g_{i'}$ for a second LOB i' the same way. Having determined g_i and $g_{i'}$, suppose now we consider all pairwise products where the first LOB is i and the second is i', without constraint on first and second companies being the same or different. We expect that the appropriate weighted average of those pairwise products will be $\sqrt{g_i g_{i'}}$. If we find this indeed to be the case, then we conclude LOBs i and i' are in the same covariance group. But if we find the weighted average to be statistically insignificantly different from zero, we conclude that LOBs i and i' are in different covariance groups. Lastly, we consider pairwise products where the first and second company is the same and where the first and second LOB is the same and equal to a selected LOB of interest. According to equation A.2, these products should display a $1/E$ dependence. Regress these products on $1/E$ and identify the intercept estimate with $c_i + g_i$. Note that c never appears naked in these expressions, always in conjunction with g, but, having already inferred g, we can back out g to infer c.

For the rest of this Appendix we will carry out the following program:

1) In the next section, "Model for Expected Losses," we will discuss the estimation of the E_{ijk} and calculation of the normalized deviations Δ_{ijk} with an adjustment for degrees of freedom. The need for weights and the appropriate weights to use in modeling E_{ijk} will be important issues.

2) The following section, "Model for Loss Variances," will discuss the use of squared normalized deviations Δ_{ijk}^2 to fit the $1/E$ variance models mentioned above and estimate the sums of contagions and covariance generators by LOB, $c_i + g_i$.

3) The following section, "Other Pairwise Products," will discuss the use of other pairwise deviation products $\Delta_{ijk}\Delta_{i'j'k}$ with at least one of i i' or j j'. Products in which the first and second LOBs are the same, $i = i'$, but companies are different, j j', yield estimators for the covariance generators g_i. Products in which the first and second LOBs are different, i i', provide a test of whether two LOBs are in the same covariance group or not. The issue of weights will again be important. Also to be introduced at this point will be the use of the bootstrap to quantify standard errors of estimates.

4) The last section, "Some Representative Results," will discuss for two lines of business some representative results for contagion c_i, covariance generator g_i, and whether or not

these two lines are in the same covariance group. Furthermore, for one of our representative lines, we will also perform the calculations measuring deviations relative to means not adjusted for long-term trends. We will indeed find much larger contagions and covariance generators. But, as we have already argued, these larger parameters are not appropriate for capital requirement calculations.

Model for Expected Losses

As already noted, we start with paid losses by LOB, by company (or company group), by year developed not to true "ultimate" but rather to the greatest age in loss development triangles available to us. We ratio these losses to premiums, build models for expected loss ratio, then multiply by premium to get back to estimates for expected loss. For each LOB, we actually test a number of denominators (premium, PPR, one or more exposure bases) in search of a denominator that produces a model for the ratio of loss to that denominator with a relatively high R^2. Presumably, for those denominators producing ratio models with lower R^2, the additional unexplained volatility is attributable to the denominator and interferes with good estimates for expected loss. High R^2 means the denominator is either stable or changes smoothly over time and is less likely to interfere with good estimates of expected loss.

Graph A.1.1 shows loss ratios by year, each line representing a separate company or company group. This is a package line with considerable property exposure, which may explain the apparent coordinated short-term up and down movement, which is evidence of correlation across company within LOB. The long-term apparent upward trend is probably just that, trend, was probably predictable, and, according to the discussion at the beginning of this Appendix, should not be considered evidence of correlation in the sense that we mean correlation.

Graph A.1.2 shows loss ratios by year for a liability line. Correlation is less readily apparent in this second graph. We should not be surprised if the correlation parameters we estimate for the second LOB are less than those for the first.

The graphs for these two lines are reasonably representative of graphs for the other lines we studied as well. The reader should note an important feature of these graphs that motivates the subsequent model. The lines for some companies lie consistently above the lines for other companies and appear to move in parallel to one another. Where correlation is visually significant (LOB 1), the parallel motion is evident even over short periods of time. Where correlation is less visually significant (LOB 2), the parallel motion is less pronounced over short periods of time but is still evident, on average, over the decade as a whole. This suggests a main-effects model with main effects for company and year. We assume no company/year interactions partly because such interactions are not apparent on the graphs and partly because we could argue that we lack sufficient data to estimate separate year effects by company anyway. We fit the year effect with a non-linear, non-parametric smoother to capture a wide range of possible behaviors across years – consistent trend, turning points, points of inflection, etc. This model produces fitted loss-ratio values that are parallel curves, a separate curve for each company.

The fitting is performed by invoking a generalized additive model package, specifying normally distributed errors, an identity link function, main effects for company and year, and a loess smoother on year with wide smoothing window (large "span"), so as not to respond too much to random hits in any one year. Although one could argue that, technically, loss ratios cannot be normally distributed (shouldn't be negative and are likely positively skewed), we observed deviations from normality sufficiently mild for our data that the normal assumption was acceptable, which brought us that much closer to the classic linear model. Also, we saw no evidence that the loss ratios themselves were not additive in the explanatory variables (company and year), hence the identity link function, which again brings us that much closer to the classic linear model. In fact, the only reason for invoking the generalized additive model, rather than the classic linear model, was our desire to impose a non-linear, non-parametric smoother on the year effect.

The generalized additive model was weighted. Over the years, it has been our experience fitting statistical models to insurance data that unweighted models are almost never appropriate. Weighted models are generally more appropriate, because insurance data points are almost never of equal credibility or volatility; and, furthermore, the range of credibilities or volatilities is sufficiently great that unweighted models are inadvisable. The general statistical practice is that the weight associated with a data point varies as the reciprocal of its variance. This practice produces minimum-variance fitted values. A general statistical rule of thumb is that, so long as the variances of the data points are sufficiently similar to one another (in other words, differ from one another by no more than a factor of two or three) and assuming the variances independent of the explanatory variables in the model, then the differences in results between a weighted and an unweighted model can be expected to be sufficiently modest that they are ignorable. Then an unweighted model is acceptable. The purpose of weighting is not to adjust for every last bit of difference in variance but rather to correct for gross asymmetries in variance. But most insurance data presents a range of variances considerably greater than a factor of two or three and so generally calls for the estimation of weighted models.

The classic actuarial assumption is that the variance of a loss ratio declines as one over some measure of volume, such as premium, which would suggest weighting on premium. But the formulas of the previous section of this Appendix would suggest that, in the presence of parameter uncertainty, the variance depends on two terms, one of form $1/volume$, the second a constant greater than zero. So the very smallest risks, for which the $1/volume$ term dwarfs the constant, do indeed see a variance declining as $1/volume$. The very largest risks, for which the $1/volume$ term has essentially died away to zero, see a variance essentially independent of size. If all the data is essentially small risks, weighting on volume is appropriate. If all the data is essentially large risks, doing an unweighted analysis is reasonable. Generally, we are somewhere in the middle, with risks all the way from the small to the large.

One possibility is to construct an iterated model. Select some weights. Fit a weighted model to find fitted means. Find the differences of actuals and fitted means, square the differences, and fit these squared differences to the variance model $1/volume$ plus a constant. Invert the fitted variances to find a new set of weights and iterate a few times. This is admittedly a fair amount of work. A "quick and dirty" alternative that we have frequently found to work adequately for weighting, where adequate means it removes gross asymmetries in variance without necessarily reducing all variances to exact equality, is to assume that variance dies away as 1 over some fractional power of volume; say, variance dies away as $1/\sqrt{volume}$ --hence use the square roots of volumes as weights. Over quite a robust range of different models, we have found that this square root rule roughly captures the change in volatility from the small to the large.

As an example, Graph A.2 shows the same loss ratios as in Graph A.1.1 (LOB1), but plotted against premium rather than year. The smallest risks have premium as small as approximately $5 million. The largest premiums exceed $1 billion. So premium covers a range of two and a half orders of magnitude. As expected, loss ratio volatility appears to decline with increasing volume, but apparently not as fast as a 1/*volume* rule would imply. If the 1/*volume* rule held, as premium increased by more than a factor of 100, variances on the extreme right would be less than 1/100 of the variances on the extreme left, and standard deviations on the extreme right would be less than 1/10 of standard deviations on the extreme left. Standard deviations on the extreme left don't look 10 times as big as standard deviations on the extreme right--more like the three or four times as big that would be implied by variances that went as $1/\sqrt{volume}$; hence standard deviations that went as $1/\sqrt[4]{volume}$. So, in building our models for loss ratio for LOB 1, we have used weights of $\sqrt{premium}$. In other words, data points associated with the largest risks are assigned weights on the order of 10 times as large as data points associated with the smallest risks.

Graph A.3 shows the year effect for this model on LOB 1. The dotted lines are the fitted year effect plus and minus two standard errors, corresponding to an approximately 95% confidence interval. The year effect has been translated to yield an average effect of 0. The absolute level of loss ratios is captured by the other main effect, the company effect. So we see loss ratios have been trending upwards throughout the decade, increasing by more than 20 loss ratio points from the beginning to the end of the decade, but the trend has not been uniform throughout. There is a point of inflection at mid decade. Throughout the first half-decade, trend was positive but decreasing, until it vanished altogether at mid-decade, only to resume its upward movement at decade end. Because this happened to all companies (at least our model assumes so, being a main-effects-only model, but, as noted before, there is no evidence of different year effects by company), and because the trend was essentially consistently upward and of significant magnitude, if we were to measure deviations about the decade mean, we would find most deviations early in the decade negative, most late in the decade positive. We would infer considerably larger correlations from these deviations than from deviations measured about the varying-year effect plotted in Graph A.3. For illustrative purposes only, we have actually done both calculations and will report the results later in this Appendix.

This year effect has a cubic appearance. This shows the importance of the non-parametric component of the smoother on year. Because the smoother was locally quadratic, in the absence of a non-parametric component, the global year effect would have been linear or quadratic and could not have captured the pattern evidenced in Graph A.3. At the same time, the smoother is not so responsive as to pick up the year-to-year ups and downs apparent in Graph A.1.1. So long-term trends captured in the means, as driven by the year effect, therefore are removed from deviations about means, and don't impact correlation estimates. Short-term ups and downs are not captured in the year effect or the resulting means, so do flow through to deviations about those means and do carry through to correlations. This is the desired behavior.

Having identified good models for ratio of loss to one of premium, PPR, or exposure, we multiply the fitted values resulting from these models by the denominators to yield estimates for mean losses. These mean losses are then used to calculate the normalized deviations of the previous section of this appendix. As noted in the previous section, the normalized deviations are the actual loss minus the expected loss, the difference then divided by expected loss.

There is one additional, important adjustment to the normalized deviations not already discussed. These deviations are adjusted for degrees of freedom by multiplying by $\sqrt{n/(n-p)}$, where $n, p,$ and the justification for this particular multiplier will now be described. Suppose the model for loss ratios for a particular LOB is based on n observed data points. The fitted model has p effective parameters, where p is the number of companies, plus two (because of the locally quadratic nature of the year smoother), plus the additional effective number of degrees of freedom of the non-parametric component of the year smoother, which was generally in the neighborhood of 0.8. An unbiased estimator for variance involves taking differences of actual and fitted values, squaring the differences, summing up the n squared differences, and dividing the sum not by n but by $n-p$. The way in which we subsequently use the normalized deviations to estimate correlation parameters amounts to taking averages, dividing sums of n terms by n rather than by $n-p$. By adjusting normalized deviations by the factor $\sqrt{n/(n-p)}$, we are adjusting squared deviations by $n/(n-p)$, the n's cancel, yielding the right denominator, $n-p$, in the end.

The need for applying a multiplier greater than 1 to the unadjusted normalized deviations can also be seen from the following argument, although this argument doesn't also establish the magnitude of the multiplier. We start with *n* data points. To these data points we fit a model with *p* effective degrees of freedom. The fitted values are themselves random variables that approximate the "true" expected values to the extent that the model is the "true" model. But note that fitted values are pulled in the direction of the observed data and away from the true expected values by the fitting process (least squares, maximum likelihood, whatever). The magnitude of differences between actual and fitted values will therefore be smaller on average than the magnitude of differences between actual and true expected values. This shrinkage can be offset by multiplying the first differences by $\sqrt{n/(n-p)}$, where the actual value of the multiplier is established by the requirement that sums of squares reproduce the right unbiased estimate for the variance.

In the interests of wrapping up loose ends, we should note that, although we always started with a model with main effects for company and for year, with a smoother for year, the finally accepted models were many different variants on this. We sometimes found that company was not statistically significant; in other words, there was no statistically significant evidence that loss ratio differed by company. We sometimes found that the non-parametric component of the year effect was not significant, so the year effect was globally quadratic. Sometimes the quadratic term was not significant, so the year effect was globally linear (long-term constant trend). And sometimes even the linear effect was not significant, so there was no statistically significant evidence of loss ratio varying across years at all.

Model for Loss Variances

So now we have normalized deviations, adjusted for degrees of freedom. We consider all manner of pairwise products of these deviations. We demand that the year associated with the first factor in the pair match the year associated with the second factor, because we have not yet studied correlations across year. If we consider just those pairwise products where the first and second company also match, and where first and second LOB also match and are equal to some specified LOB of interest, then we are looking at squared deviations. Equation A.2 suggests that, if we plot these squared deviations against expected loss *E*, we should see a *1/E* dependence plus a constant term, where the constant is the contagion plus the covariance generator for that LOB. See Graph A.4 for the graph just described for LOB 1. The circles represent the squared deviations from data. The triangles are the fitted values of the functional form $1/E$ plus constant.

The fit was created by least squares regression. There is again an issue of weights. Squared deviations for small expected loss appear considerably more volatile than squared deviations for large expected loss, and so should receive less weight. Otherwise, there is a considerable risk that some noisy data at small E could have a considerable impact on the estimate of the constant term out at large E. What weights might be appropriate? If the deviation Δ were approximately normal with standard deviation σ, then Δ^2/σ^2 would be distributed approximately chi-squared with one degree of freedom. This result would imply that Δ^2 has an expectation of σ^2 and a variance of $2\sigma^4$. In other words, the standard deviations of the squared-deviation random variables appear proportional to their expected values, which is not inconsistent with Graph A.4. This suggests the following algorithm. Fit the $1/E$ plus constant functional form to the squared deviations. Square the fitted values, take their reciprocals, and use these values as weights in another fit of the functional form to the squared deviations. Iterate a few times.

Other Pairwise Products

Consider next pairwise products where first and second year are the same, first and second LOB are the same and equal to some specified LOB of interest, but first and second company are different. These products measure correlation among companies within LOB, and their (weighted) average yields an estimator for the covariance generator for that LOB, per equation A.2. Consider first a plot of the second factor in each pair against the first factor in each pair. Can one visually see the correlation? Graph A.5.1 is such a plot for LOB 1.

The most striking thing about this plot is that the data appears to array itself in rows and columns. Consider an example. Suppose for this LOB we have 10 years, 10 companies, hence 100 independent observations from which we construct 100 normalized deviations. For each of the 100 deviations thought of as the first factor, there are nine deviations available as second factor (same year, each of the other nine companies), hence a total of 900 pairwise products relevant to this section of the Appendix (same year, different companies) and 900 plotted points on the plot of second factor vs. first factor of the form of Graph A.5.1. The points in this plot array themselves in columns of nine points and rows of nine points. The columns of nine result because all nine share the same first factor (plotted on the x axis) while the second factor (plotted on the y axis) ranges over nine possible values. Rows of nine also result because all nine share the same second factor while the first factor ranges over nine possible values. The nine points in a column are not independent but highly interdependent through their shared first factor. Likewise, the nine points in a row are not independent but highly interdependent through their shared second factor. These interdependencies through shared first and second factors apply also to the 900 pairwise products. It would be very wrong to treat these 900 pairwise products as 900 independent draws from some underlying process. This observation will be relevant to a later discussion of standard errors of parameter estimates, such as estimates of covariance generators.

Returning to Graph A.5.1, note the slightly tilted horizontal line. This is an unweighted linear regression line on the plotted points. It is included as an aid to visualizing a possible tilt to the plot, which would be indicative of a correlation, but the degree of tilt of this regression line is not a good estimator of the correlation. First, points with either very low or very high first deviation may be highly leveraged and highly influential in estimating the unweighted regression line. Yet these extreme first deviations are likely to be the most volatile and the least deserving of receiving any significant weight. An unweighted regression gives them too much weight. Second, the regression line treats all the plotted points as independent of one another, and we have already argued that there is a great deal of interdependency among these points. So the plotted regression line should be treated as a visual aid only and not considered a good estimator. We have argued in a previous section of this Appendix that a weighted average of pairwise products, with judicious choice of weights, might be a good estimator of covariance generators.

The deviations of Graph A.5.1 are those measured about expected losses taking into account the year effect of Graph A.3. As an additional aside on the potential distortion of estimating correlations from deviations about grand means, Graph A.5.2 shows a plot corresponding to Graph A.5.1 of deviations vs. deviations, measured about expectations not reflective of the year effect. The apparent correlation is much greater, the excess correlation being driven by the failure to remove long term predictable trend from the deviations.

We have concluded that, because of various technical difficulties, plots of deviations vs. deviations of the form of Graph A.5 are useful visual aids but not good estimators. As weighted averages of pairwise products of deviations can be used as estimators, what weights are appropriate? Previously, we presented a heuristic argument in terms of the chi-squared distribution for squared deviations; in other words, for pairwise products where the first and second factors are identical. But we don't know what the sampling distribution might be for pairwise products of deviations where the first and second factors may be interdependent but not identical. Suppose we plot pairwise products against some measure of volume to see if there is any evidence of changing volatility with increasing volume. For each of the first and second factors of a pairwise product, there is a measure of volume, namely the expected loss associated with that deviation, but the two expected losses are unlikely to be equal. Suppose we define as a measure of volume for the pairwise product the geometric average of the expected losses for the first and second deviations in the product; in other words, the square root of the product of the two expected losses. Call it E.

Graph A.6.1 shows a plot for LOB 1 of the pairwise deviation products, same year first and second factors, different companies, against this volume measure E. Pairwise products associated with larger volumes are clearly less volatile and so should receive more weight in any weighted average of these products. Suppose we imagine that the variance of the sampling distribution of a pairwise product declines as unity over some power of E. Dividing the observed pairwise products by the square root of the presumed variance law and plotting this against E should produce a graph more symmetrical left to right than Graph A.6.1. Suppose we guess the variance law to be $1/E$. Then multiply pairwise products by $\sqrt{E}$. Graph A.6.2 shows this plot. We have gone from a graph that shows more volatility on the left to one that shows more volatility on the right. Clearly, a $1/E$ variance law overdoes it. Suppose we assume a variance law $1/\sqrt{E}$. Then multiply pairwise products by the fourth root of E. Graph A.6.3 shows the resulting plot is far more symmetric than either A.6.1 or A.6.2, supporting a variance law something like $1/\sqrt{E}$ and, therefore, a weighted average of pairwise deviation products with weights proportional to $\sqrt{E}$ as a reasonably best estimator from among this family of estimators of the covariance generator for this LOB.

Now that we have an estimate for the covariance generator, how precise is it? What is the standard error of that estimate? Generally, when an estimator is a weighted average of independent observations, the standard error of the estimate is the standard deviation of one observation divided by the square root of the number of observations, with some adjustment for the weighting. As we have already argued, these pairwise products are far from independent of one another, ruling out the square root of *n* rule. We have chosen to estimate standard errors of estimators via bootstrap. From the original data draw a data resample of the same size as the original data set, but with replacement, so that some data points might not appear at all in the resample and others might appear more than once. Re-estimate the statistic or parameter of interest from this resample. Repeat this many times, building up a collection of estimates, from which collection one can estimate such quantities as the standard deviation and extreme percentiles of the estimator. Statistical rules of thumb suggest that, whereas one may need hundreds of resamples to reasonably estimate extreme percentiles (such as the 95th or 99th) of the sampling distribution of the estimator of interest, as few as fifty resamples will yield a reasonable estimate of the standard error of the estimator.

Furthermore, to preserve the two-way structure of the underlying problem on company and year, as well as to estimate the relative impact of company and year on estimators, we bootstrap separately on company and year. Bootstrapping on company yields a standard error of the estimator due to the randomness of which companies are in or out of the database. In other words, if certain companies were dropped from the database, and certain others were added, how much could we expect the estimator to vary from its current value? Bootstrapping on year yields a standard error of the estimator due to the randomness of which years are in or out of the database. The total standard error of the estimator is the square root of the sum of squared standard errors due to company and year separately.

An example may again be useful. Suppose our previous example with an LOB with ten years and ten companies. This produces 100 normalized deviations, 100 squared deviations used to estimate the variance model, and 900 pairwise deviation products, first and second years the same, first and second LOBs the same and equal to the LOB in question, but different first and second companies, from which an estimate for the LOB covariance generator is calculated. One way to bootstrap would be to draw from the 100 deviations with replacement, but it is likely that this would produce a resampled dataset in which some years were represented by some companies but not all ten companies, and some companies were represented by some years but not all ten years. The resampled dataset would not preserve the two-way structure of the original on company and year. Also, from this resample it would be impossible to segregate the potentially interesting different impacts of company and year.

We chose to resample on company and year separately. One resamples on company by drawing ten companies with replacement from the original list of ten. As an example, the resampled list might include eight of the original ten appearing once each, the ninth appearing twice, and the tenth not at all. Then one takes all ten years for each of the resampled companies. The result would be 100 deviations, the first 80 from the original 100 representing the first eight companies, then 81 through 90 from the original 100 representing the ninth company, then 91 through 100 repeating 81 through 90, representing the ninth company showing up a second time in that particular resampling on company. So, although the resample includes 100 deviations, there are only 90 distinct values, because company 9 occurs twice in the resample. One uses these resampled 100 deviations to calculate the previously discussed variance model and covariance generator estimator. Resample 50 times to estimate standard errors for the estimators.

Next resample on year by drawing ten years with replacement from the original list of ten. As an example, the resampled list might include six of the original ten appearing once each, the seventh and eighth appearing twice each, and the ninth and tenth appearing not at all. Then take all ten companies for each of the resampled years. The result would be 100 deviations but only 80 distinct values, because years 7 and 8 occur twice in the resample. Use these resampled 100 deviations to calculate the previously discussed variance model and covariance generator estimator. Resample 50 times to estimate standard errors for the estimators.

The previous section of this Appendix, on the variance model, considered pairwise deviation products where the first and second factor years were the same, first and second LOBs the same, and first and second companies the same; in other words, the pairwise products were actually squared deviations. These lead to variance models and estimators for the sum of contagion and covariance generator for the LOB. In this section, we have considered pairwise products with first and second years the same, first and second LOBs the same, but first and second companies different. These products lead to estimates of correlation among companies within LOB, to estimators for the LOB covariance generator. Other pairwise products not yet discussed but of potential interest would be those for which first and second years are the same, but first and second LOBs are different. Such products would lead to estimates of between-LOB-correlation, to estimators for the geometric average of the covariance generators for the two LOBs if they are in the same covariance group, or to a statistic not statistically different from zero if the LOBs are in different covariance groups. We will not discuss these products further other than to note that the weighting and bootstrap issues discussed above are the same for these products and were addressed in the same way.

Some Representative Results

Before discussing Exhibits A.1 through A.3, which provide some representative results, we should note that we tested two other model issues that have not yet been discussed.

1) Between company pairwise deviation products yield estimators for covariance generators. We asked whether there was any evidence that these covariance generators varied by size of company. We tested this by regressing the appropriate pairwise products against the base 10 logarithm of the size of the company, size measured as the expected loss for that LOB. A statistically significant regression coefficient for the log explanatory variable would have been evidence of a size dependency. A statistically significant positive coefficient would have been evidence of a covariance generator increasing with increasing company size, and vice versa for a statistically significant negative coefficient. We used log(size) as the explanatory variable on the assumption that the effect, if there was one, would be logarithmic in size, that the magnitude of the effect would be about the same when going from a company of size 1 to size 10 as when going from a company of size 10 to one size 100, etc. No statistically significant size effects for the covariance generators were detected.

2) For certain property lines, we asked whether much of the apparent correlation arose through catastrophes. We eliminated the heavy catastrophe years of 1992 and 1994 and found that correlations did indeed go down but were still significant.

Turning now to Exhibit A.1, this exhibit considers just pairwise deviation products where first and second LOB are LOB 1. Considered first are products where first and second companies are different ("Between companies"), hence the expectation is g_1. Based on a weighted average of the relevant pairwise products from the data, the point estimate for g_1 is 0.0026. The square root of this, 0.051, is the standard deviation of the underlying frequency multiplier, which appears to indicate a frequency parameter uncertainty impacting LOB 1 industry wide of on the order of plus or minus 5%. Bootstrapping on years yields a range of estimates for g_1 with a standard deviation of 0.0008. Bootstrapping on companies yields a standard error due to companies of 0.0009. So uncertainty regarding this parameter due to years is comparable to the uncertainty arising through companies. The total standard error for g_1 is a combination of standard errors due to years and companies and is 0.0012. The estimate for g_1 is more than twice its standard error, so is certainly statistically significant.

The test for g_1 size dependence yields a regression coefficient for the *log*(*size*) explanatory variable of -0.00004, with a standard error estimated from bootstrap of 0.00344. The standard error is much larger than the parameter estimate. There is no statistically significant evidence that g_1 depends upon size.

Considering next pairwise products with first and second LOB equal to LOB 1 and with first and second companies equal ("within company"; in other words, the squared deviation products) yields an estimate for LOB 1 of contagion plus covariance generator of 0.0226 with a standard error of 0.0092. This is certainly significant. The difference of the $c + g$ estimate (0.0226) and the g estimate (0.0026) yields an estimate for the contagion c for LOB 1 of 0.0200.

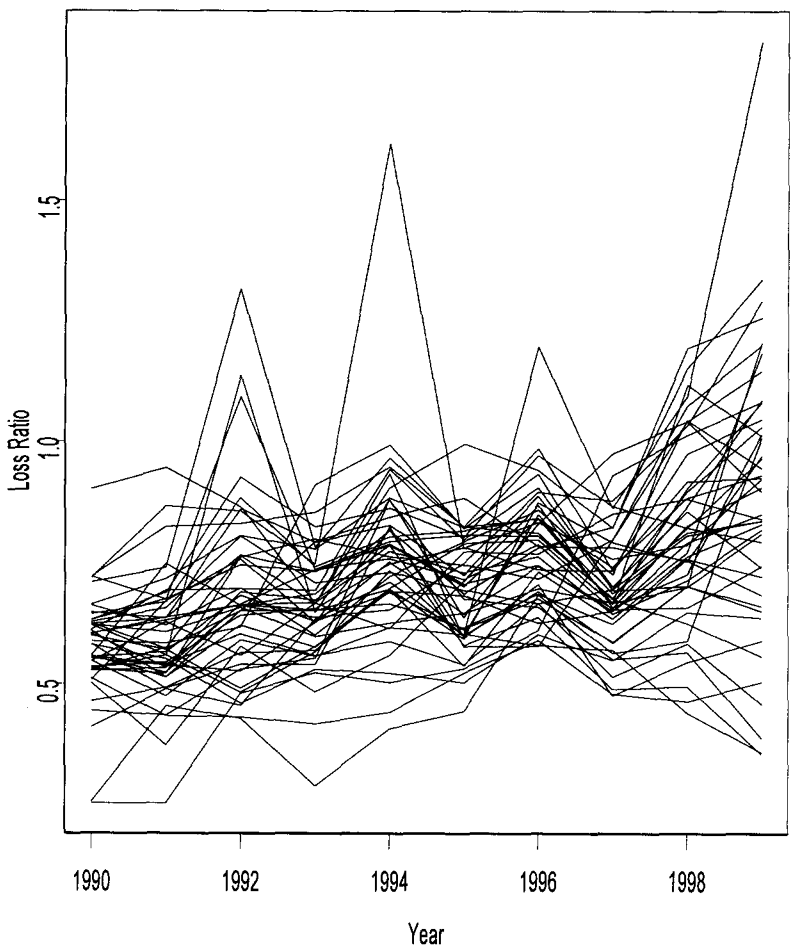
If, just for the sake of illustration, not that we argue this is the right thing to do, we repeat these calculations for LOB 1 using deviations about grand means rather than about means adjusted for the year effects of Graph A.3, we find much larger correlation estimates. For g_1, instead of the 0.0026 with standard error 0.0012 discussed above, we find 0.0135 with standard error 0.0051. This latter value for g_1 implies a frequency parameter uncertainty of 11.6% vs. the 5% discussed above. Likewise, for $c_1 + g_1$, instead of the 0.0226 with standard error 0.0092 discussed above, we find 0.0298 with standard error 0.0099. Failing to adjust deviations for long-term predictable trends significantly inflates correlation estimates in ways not directly relevant to the required capital issue.

Exhibit A.2 shows the same statistics for LOB 2, a g estimate of 0.0007 with standard error of 0.0004 (hence just about significant at two standard errors, indicating a frequency parameter uncertainty of plus or minus 2.6%), no significant size dependence of this g estimate, and a significant estimate of $c + g$ of 0.0090 with standard error of 0.0023. From comparing Graphs A.1.1 and A.1.2 we had suspected we would find more correlation in LOB 1 than in 2, and indeed we find g for LOB 1 larger than that for LOB 2. $c + g$ measures large risk volatility (the limit as the $1/E$ term dies away). This is also larger for LOB 1 than for LOB 2.

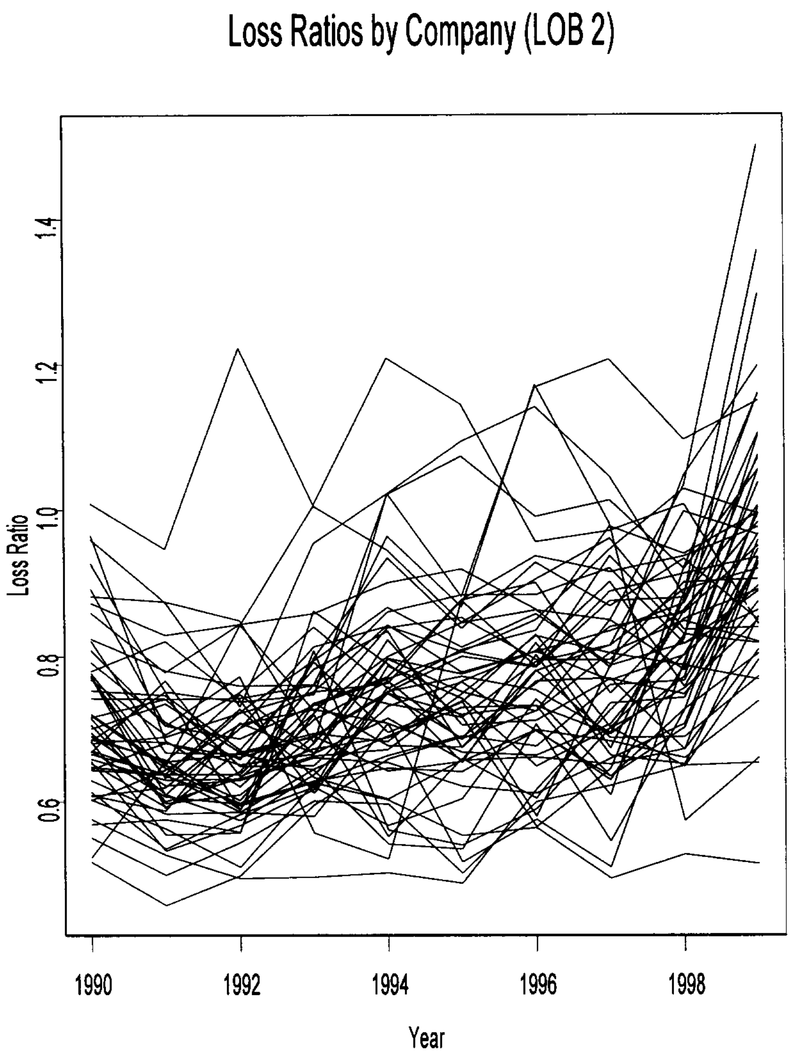
Turning lastly to Exhibit A.3, this considers pairwise products where the first LOB is LOB 1 and the second LOB is LOB 2, hence measures between LOB correlations. This yields an estimate of $\sqrt{g_1 g_2}$ of 0.0005 with a standard error of 0.0006. Because this statistic is not statistically significantly different from 0, there is no evidence that LOBs 1 and 2 are in the same covariance group. Knowing what lines of business LOB 1 and 2 are, we did not expect them to be in the same covariance group and are not surprised by this result.

Graph A.1.1

Loss Ratios by Company (LOB 1)

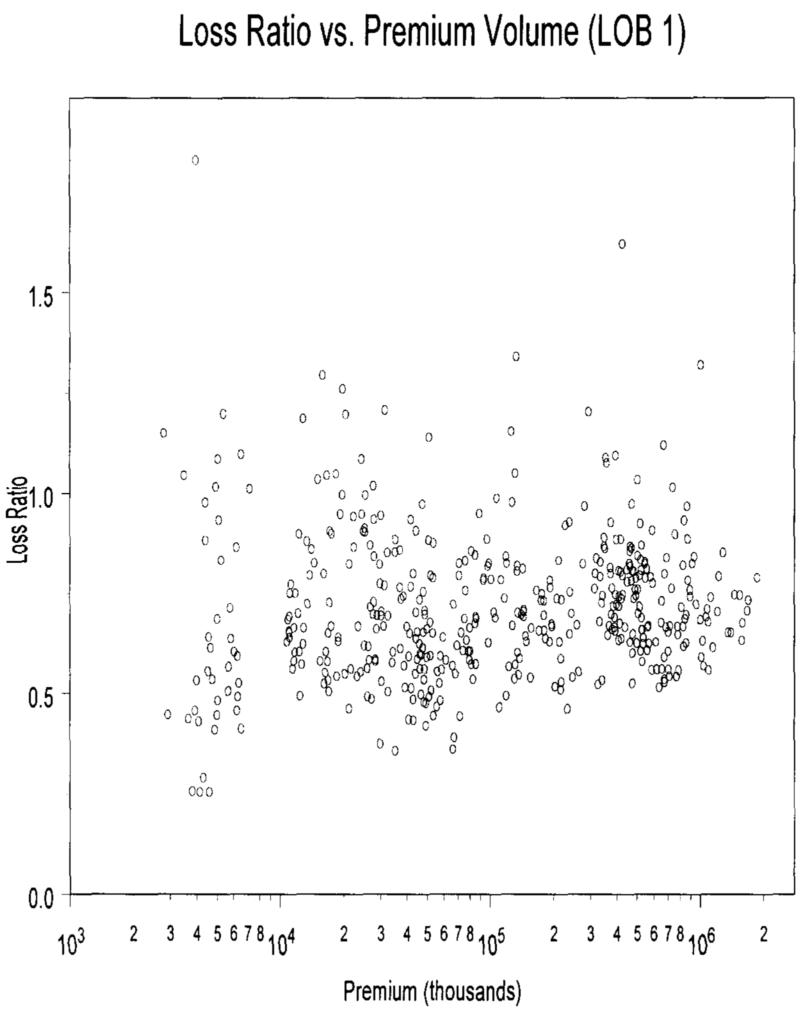


Graph A.1.2



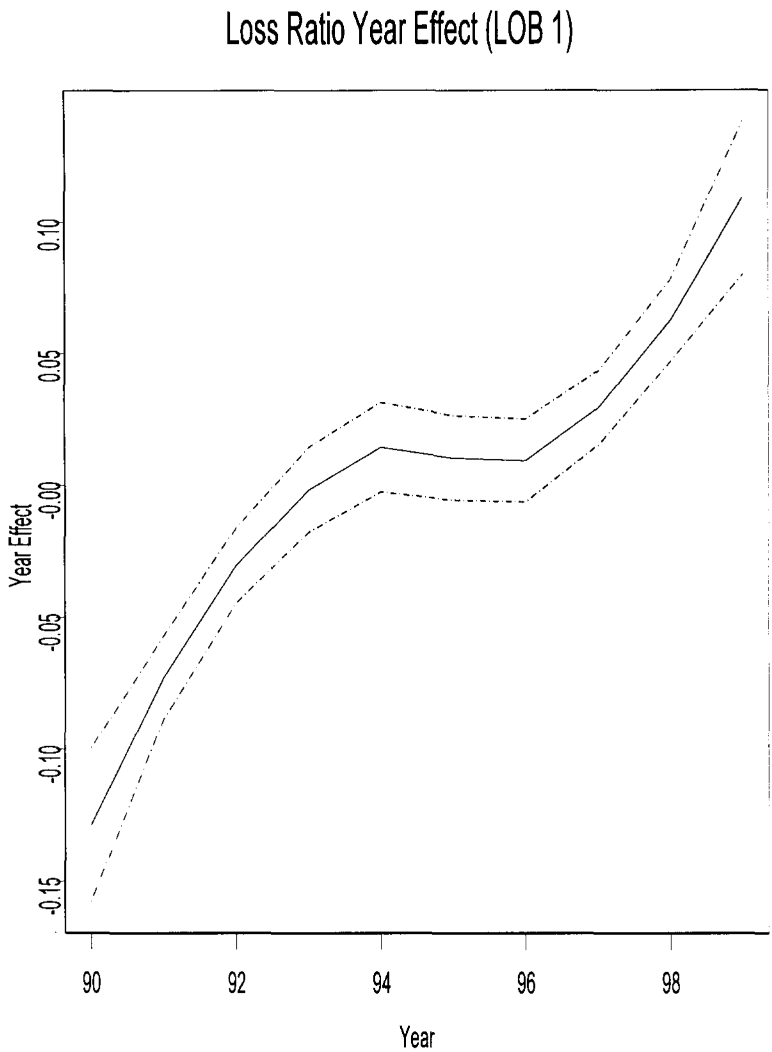
Loss Ratios by Company (LOB 2)
Loss Ratio
0.6
0.8
1.0
1.2
1.4
1990
1992
1994
1996
1998
Year

Graph A.2



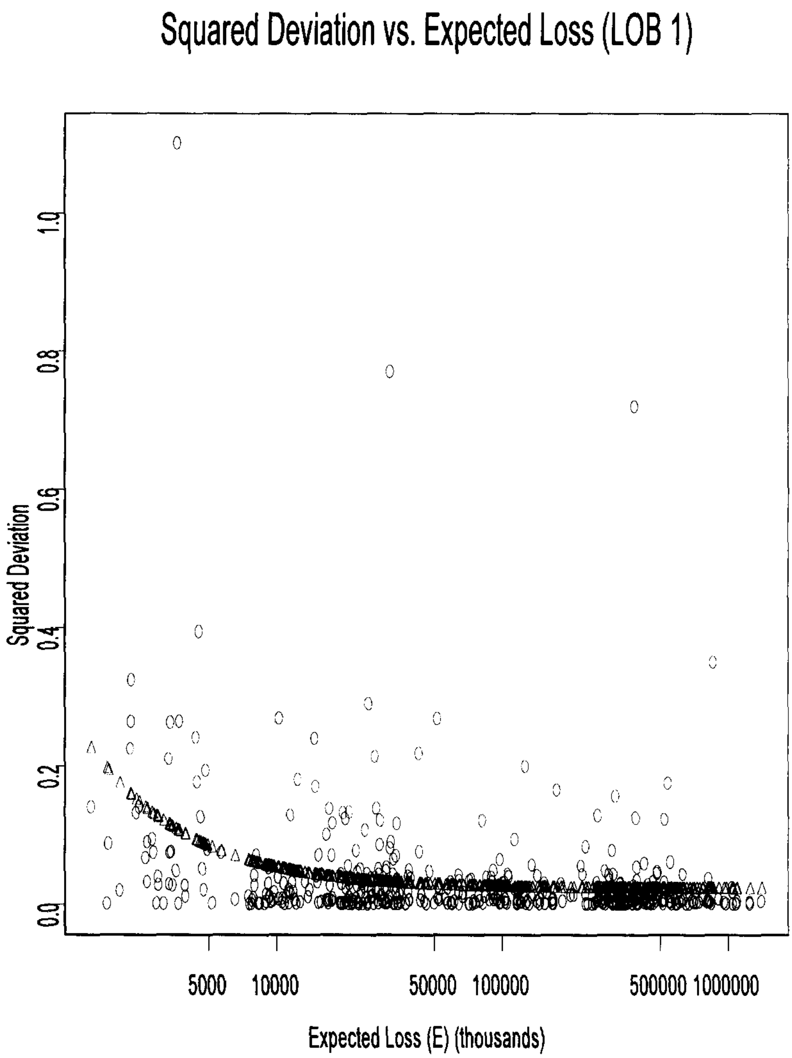
Loss Ratio vs. Premium Volume (LOB 1)
Loss Ratio
1.5
1.0
0.5
0.0
10^3
2
3
4
5
6
7
8
10^4
10^5
10^6
Premium (thousands)

Graph A.3



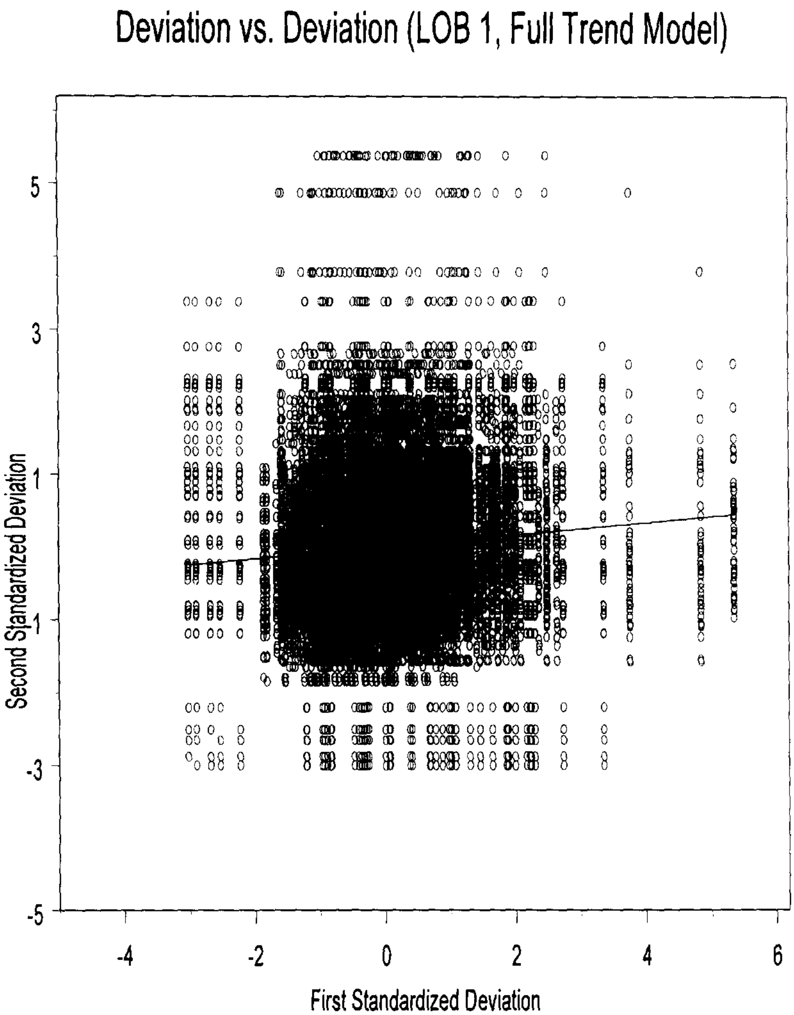
Loss Ratio Year Effect (LOB 1)
Year Effect
0.10
0.05
-0.00
-0.05
-0.10
-0.15
90
92
94
96
98
Year

Graph A.4

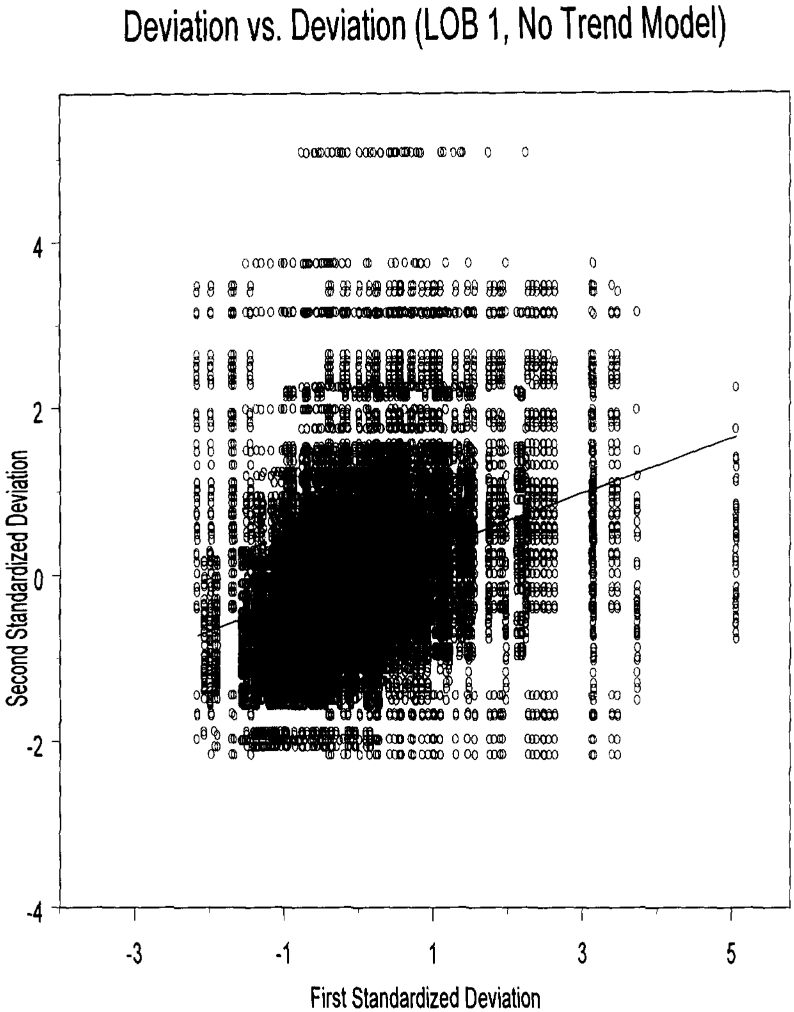


Squared Deviation vs. Expected Loss (LOB 1)
Squared Deviation
0.0
0.2
0.4
0.6
0.8
1.0
5000
10000
50000
100000
500000
1000000
Expected Loss (E) (thousands)

Graph A.5.1

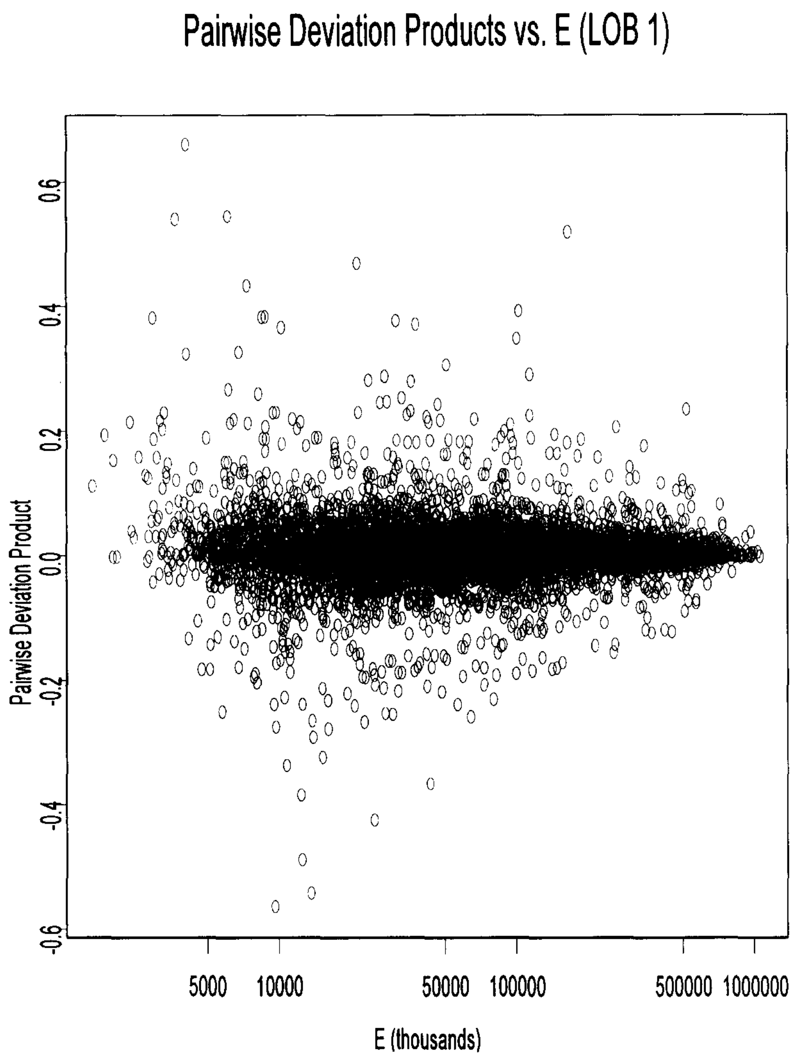

Deviation vs. Deviation (LOB 1, Full Trend Model)
5
3
1
-1
-3
-5
Second Standardized Deviation
-4
-2
0
2
4
6
First Standardized Deviation

Graph A.5.2

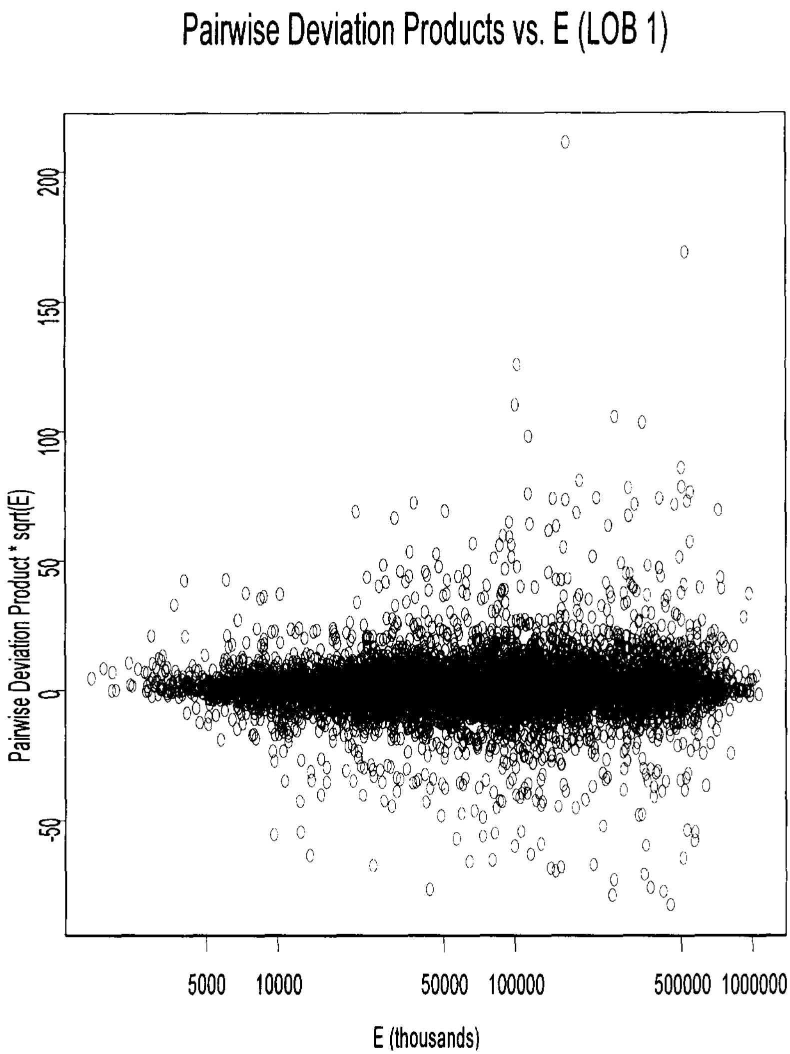


Deviation vs. Deviation (LOB 1, No Trend Model)
4
2
0
-2
-4
Second Standardized Deviation
-3
-1
1
3
5
First Standardized Deviation

Graph A.6.1


Pairwise Deviation Products vs. E (LOB 1)
Pairwise Deviation Product
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
5000
10000
50000
100000
500000
1000000
E (thousands)

Graph A.6.2


Pairwise Deviation Products vs. E (LOB 1)
Pairwise Deviation Product * sqrt(E)
-50
0
50
100
150
200
5000
10000
50000
100000
500000
1000000
E (thousands)

Graph A.6.3

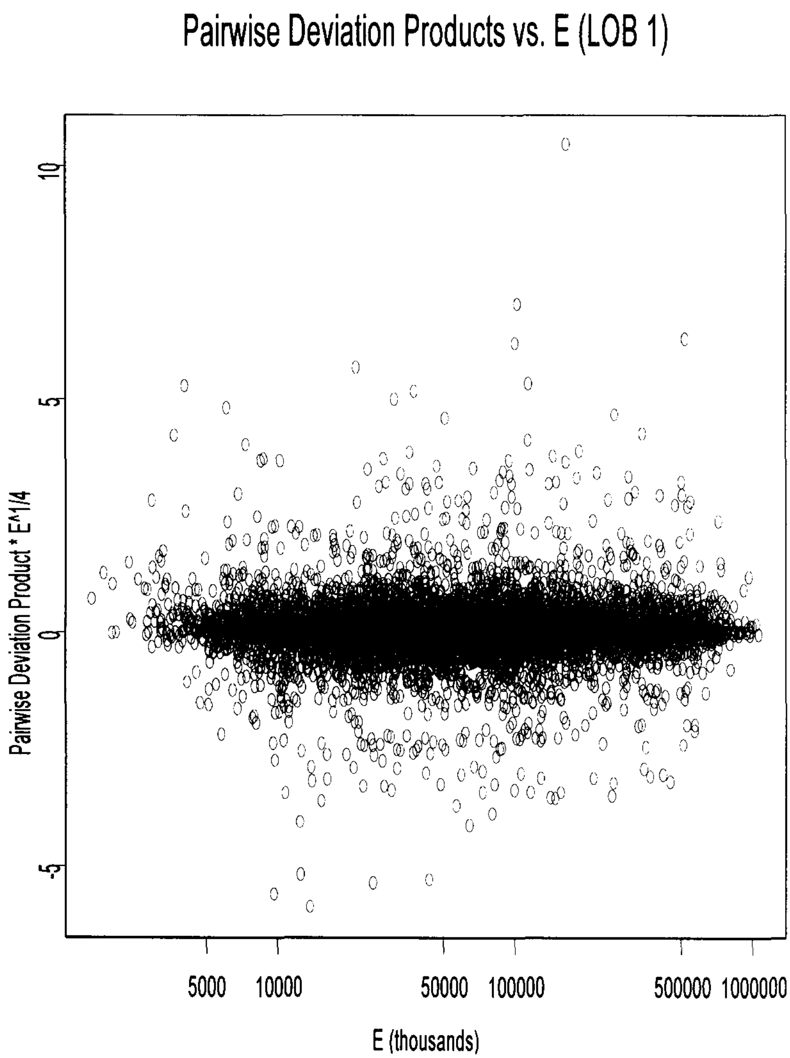

Pairwise Deviation Products vs. E (LOB 1)
Pairwise Deviation Product * E^1/4
10
5
0
-5
5000
10000
50000
100000
500000
1000000
E (thousands)

Exhibit A.1

Correlation Parameter Estimates

LOB 1

Between companies: *g*

Estimate: 0.0026

Standard error due to years: 0.0008

Standard error due to companies: 0.0009

Full standard error: 0.0012

Between companies: $\log_{10}$(size) coefficient

Estimate: -4e-005

Standard error due to years: 0.00235

Standard error due to companies: 0.00251

Full standard error: 0.00344

Within company: $c + g$

Estimate: 0.0226

Standard error due to years: 0.0048

Standard error due to companies: 0.0078

Full standard error: 0.0092

Exhibit A.2

Correlation Parameter Estimates

LOB 2

Between companies: g

Estimate: 0.0007

Standard error due to years: 0.0002

Standard error due to companies: 0.0003

Full standard error: 0.0004

Between companies: $\log_{10}$(size) coefficient

Estimate: -0.00065

Standard error due to years: 0.00050

Standard error due to companies: 0.00065

Full standard error: 0.00082

Within company: $c + g$

Estimate: 0.0090

Standard error due to years: 0.0007

Standard error due to companies: 0.0022

Full standard error: 0.0023

Exhibit A.3

Correlation Parameter Estimates

LOB 1 vs. LOB 2

Between and within companies: *g*

Estimate: 0.0005

Standard error due to years: 0.0005

Standard error due to companies: 0.0003

Full standard error: 0.0006

Between and within companies: $\log_{10}$(size) coefficient

Estimate: -0.00086

Standard error due to years: 0.00080

Standard error due to companies: 0.00106

Full standard error: 0.00133

Advanced Modeling, Visualization, and Data Mining Techniques for a New Risk Landscpape

Lee Smith, FCAS, MAAA, and
Lilli Segre-Tossani

Casualty Actuarial Society Call for Papers
2003 Risk and Capital Management Seminar • July 28-29, 2003

ERM & DFA Modeling:
Risk Correlation, Integration, Dependency and Concentration

Advanced Modeling, Visualization and Data Mining Techniques for a New Risk Landscape

Submitted by Lee Smith and Lilli Segre Tossani

Abstract

The risk landscape that confronts financial institutions in the 21st century presents an unprecedented departure from past experience. Traditional mathematical tools based in linear statistical mathematics are failing actuarial science by not being able to deliver credible analyses in an environment characterized by issues such as multiple correlations, extreme events and cascading risks.

Complexity science, an area of scientific investigation that has been largely limited to powerful laboratory environments until recent decades, offers new tools and methodologies with which to address this new environment. *InsuranceWorld©* is an adaptive agent-based simulator designed through application of complexity science to allow an individual company to see how markets will evolve under various catastrophe scenarios. This paper describes this tool and its scientific and mathematical foundations.

Navigating the Global Risk Landscape

Contemporary global financial institutions are facing a risk landscape of unprecedented peril. Events for which there is no credible historical data, correlations that were previously undetected, products of increasing complexity and a rapidly changing information universe: all these lead to the need for tools of unprecedented sophistication. Terrorist behavior, in particular, is non-linear by nature, and not amenable to traditional modeling techniques.

Traditionally, actuaries have been able to use linear[1] statistical models to analyze the problems in their domain. The classical case is in pricing life insurance, where large historical databases have been built. The Law of Large Numbers allows for convergence of results as observations increase, given assumptions about the nature of the universe from which the data comes.

---

[1] Throughout this paper, the authors use the word "linear" in its scientific sense; that is to say, as implying parallelism between the magnitude of a cause and the magnitude of its effect. One of the characteristics of complex systems is that they are non-linear. Non-linear systems exhibit unpredictable but non-random cause and effect relationships. Edward Lorenz' "butterfly effect" provided an illustration of this behavior by postulating that a butterfly flapping its wings in Brazil might cause a tornado in Kansas. In complex systems, even a very small change in initial conditions can rapidly lead to changes in the behavior of the system that appear counter-intuitive in both nature and magnitude.

Statistical modeling has been at the heart of risk management for the past century. Most actuarial methodologies rely on the availability of a reasonably credible amount of historical data. Future activity is then estimated by evaluating historical patterns and determining how they should be adjusted to give a reasonable assessment of future patterns.

Data requirements are even greater when risk classification calculations are required. It is one thing to develop needed premium levels for an entire line of business, but then the rates need to be determined for various categories of risk. There is a traditional trade-off between what is known as homogeneity and credibility, where the actuary must decide how finely to divide up the databases available.

Actuarial Modeling

Actuaries develop elaborate modeling formulae or programs that are applied to historical data and resolve to describe a probable future to which a current dollar price tag can be attached. One reason that there is an almost infinite variety of actuarial models is that each model necessarily incorporates an element of judgment or intuition or speculation in the definition and weighting of probable future events. As if that weren't complicated enough, data must be segregated by size of loss so that policy limit and reinsurance rate calculations can be made. This is becoming an ever more difficult part of the data collection and analysis process because the changing nature of risk magnifies the effect on losses at the high end of the scale.

Actuaries work at the intersection of nature and society, contingencies and financial impacts. They are to the financial world what physicists are to the physical. Tools being used in actuarial work hold much promise in rationalizing the way public policy issues are resolved. Actuaries take aspects of mathematics, law, economics, accounting, probability, demographics, regulation, modeling and finance, and analyze the financial impact of contingent events. There is a conjunction of understanding in actuarial science, economics, and finance that that parallels that of the emerging paradigms in other areas of thought

Physics and Modeling

Physicists can make predictions about the behavior of large numbers of atoms, electrons, photons, and other small objects, but not about individual events, which are probabilistic by nature. Similarly, an actuary can make predictions about large numbers of contingencies, deaths, accidents, or catastrophes, but not about individual events or incidents.

The mathematics behind actuarial science is the same mathematics physicists use in modeling the behavior of subatomic particles and classical matter and energy. Solution of differential equations is a major component of analyzing the phenomena with which scientists and actuaries deal. Curve fitting and error analysis are key aspects of the toolbox of both professions.

One interesting analogy is the Schrodinger wave equation, which, along with Heisenberg's matrix framework, forms the mathematical foundation for key aspects of quantum mechanics. By giving a probability structure to the location of ethereal subatomic entities that eventually emerge into the matter of our world, Schrodinger's equation allows scientists to make calculations about location, energy and momentum of events in superposition.

Actuarial events can also be said to be in superposition. Until an event occurs (e.g., a terrorist attack), the actuary is estimating which of a number of possibilities of frequencies and severities may arise from a cloud of uncertainty. By using the mathematical and technological developments created by the scientific community in the past few years, actuaries can move their models forward enormously in robustness and accuracy. With its probabilistic and demographic perspective, actuarial science can model the complex nature of the physical and social worlds in new ways. This will provide a system of metrics that can incorporate the deeper knowledge that has emerged into the calculus of decision making.

Why Traditional Actuarial Tools Fail

In the past hundred years, foretelling has moved from the realm of the intuitive to the realm of the scientific, focusing on the use of mathematics and the "law of large numbers" to provide a statistical confidence that actual results would be close to expected results. For a period of time, the technology of mathematics sufficed to provide a high level of confidence in actuarial calculations. Today, however, they have demonstrated their inadequacy to the task of pricing/valuing extreme events and highly-correlated risks.

Statistical modeling has been at the heart of risk management for the past century. Industry actuaries develop elaborate modeling formulae or programs that are applied to historical data and resolve to describe a probable future to which a current dollar price tag can be attached. One reason that there are an almost infinite variety of actuarial models is that each model necessarily incorporates an element of judgment or intuition or speculation in the definition and weighting of probable future events.

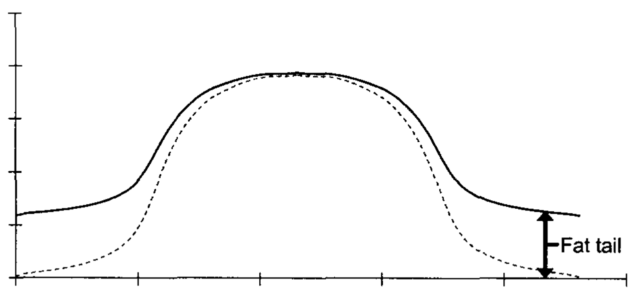
Current modeling techniques present several drawbacks. One drawback is that they do not do a good job of dealing with simultaneous changes to multiple variables in complex environments. In risk theory, for example, a portfolio distribution of risk is obtained by convoluting frequency and severity distributions. Historical data is used to determine the likelihood (frequency) of a loss arising from a given policy. (i.e., what is the likelihood this policy-holder will have one, two or x number of losses during the policy period.). Historical data is also used to determine the severity of losses (i.e., given that a loss occurs, what is the likelihood it will be $0, $1,000, $10,000, $50,000, etc.). Losses for a portfolio of risks are then estimated as the convolution of the frequency and severity distributions. So convolutions do take into account pair-wise correlations—but this is not enough. In fact, we know that in today's environment all risks are potentially cascading risks and that any individual risk will be affected by many more than one pair of correlations of varying degrees of strength.

Another drawback of current modeling techniques is that, when they are applied to large complex systems such as the national economy, there is no way to validate or test them without incurring additional, unacceptable risk. Is one, for instance, willing to force people out of their jobs in order to test the effect of unemployment on credit card delinquency?

In the past decade, the discomfort around traditional tools has escalated. We can list defining events such as Hurricane Andrew and the attack on the World Trade Center that have brought to light the potential impact of the inadequacy of traditional risk valuation and management techniques. Traditional risk-management tools were never intended to work with the expanded risks of today's global market environment.

Fat Tails

It has become clear that statistical modeling simply cannot measure the non-linear dynamics that now resonate throughout this new era of risk exposures. If there is one indicator that we can point to for validation of this failure of traditional statistical modeling, it is the emerging prevalence of "fat tails." A fat tail is a particular distortion of the classic bell curve in probability distributions. When there are fat tails, the curve looks like the solid line in the diagram below rather than the true bell curve represented by the dashed line. Complexity scientists call fat tails the signature of unrecognized correlation. Fat tails are an indicator that cascading risks are influencing the probability distribution.



Specifications for the Development of New Actuarial Tools

Thus, in the new century, actuaries are challenged daily to bring their expertise to seemingly intractable problems. A recent example is the *2002 Practice Note for Statements of Actuarial Opinion*, which directs that actuarial opinions now must consider the impact of terrorism exposure on a company's financial condition. Included in the list of things to be considered is consideration of major risk factors, coverages involved, reinsurance exposure and claim procedures.

The insurance process itself is a complex system composed of contingencies, customers, providers and financial markets all interacting to produce an insurance system that moves money based on contractual agreements. Unlike philosophers, physicists and economists, actuaries work beneath the surface of society, in arcane areas of commerce. Their tools, however, being multi-dimensional and fuzzy, may prove to be among the most useful in unraveling some of the deeper issues now being faced by society.

As the complexity and sophistication of its tools increases, actuarial science can help uncover the deep connections between nature, human behavior, institutions, finance and mathematics that drive social reality. It can help design structures by which to handle the contingencies of existence that intersect with economics and commerce.

"New" Tools Tied to Old Paradigms

Over the past decade, actuarial science has made some important attempts to incorporate new tools to meet the new challenges. The introduction of Enterprise Risk Management and Dynamic Financial Analysis into the actuarial toolbox is a good example. It may be helpful to briefly describe several of these tools in the context of the discussion that forms the topic of this paper. It is the position of the authors that these new tools, as useful as each one may be, all fall short of the need for a tool that will credibly model environments rife with multiple correlations and cascading risks.

Financial Risk Management (FRM) is both a generic term referring to the assessment and control of risk in financial markets and a very restricted professional designation. In recent years, casualty actuaries have become heavily involved in modeling phenomena on the asset side of an insurer's balance sheet as well. Financial risk management is becoming a major focus of the profession, as evidenced in contributions made in DFA and ERM modeling. The Capital Asset Pricing and Black Scholes models that are used in asset value determination use much of the same math as models used in various pure science investigations. The Brownian motion that characterizes stock price movement also arises in studies of atomic motion.

Enterprise Risk Management (ERM) uncovers the interconnectedness of operational risk, event risk, asset risk, liability risk, information risk and strategic risk. For insurance companies, it shows how the risks to the insurer itself are often correlated with the risks assumed under its coverage offerings. The major failing of ERM is that it assumes that each risk can be assessed individually and summed to all the others to reveal the enterprise total. ERM does not systematically incorporate a methodology to accurately value the effects of multiple correlations, non-linear correlations, cascading risks or positive feedback loops.

Dynamic Financial Analysis (DFA) is a tool that helps assess the uncertainty of financial results. It is a systematic approach to financial modeling that projects financial results under a variety of possible scenarios. By looking at key financial statement items as dynamic and probabilistic, it affords an opportunity to assess future financial results under a variety of scenarios. The most serious drawback to DFA models is that they are simply a compilation of traditional mathematically-calculated statistical analyses of aggregate data, iterated multiple times. Thus, not only are they linear, but they are also ponderous and require an inordinate amount of time to complete a single scenario.

The Need for a Different Approach

The inter-relationships between and among risk events and market participants have produced non-linear, cascading effects that do not follow traditional patterns or limits—and this has opened a new dimension for enterprise risk. The insurance process, with its multiple levels of interaction, demonstrates all the characteristics of a complex system. Traditional modeling techniques are inherently poor predictors of the behavior of complex systems, and so no refinement or expansion of traditional techniques will be able to meet the needs of a complex system.

Recognizing that the insurance market is a complex system, actuarial science must now seek out models specifically designed to give information about the behavior of a complex system. Complexity science uses the emerging power of computer technology and mathematics to allow scientists to take data from society and nature and let the data create its own categories. This allows scientists to approach reality before constraining it to the categories we have created for it.

Modeling Complex Systems

There is an elemental incompatibility between the underlying assumptions of traditional modeling techniques and the nature of complex systems. Luckily, there are new data extraction, analytical modeling, data visualization, and algorithmic and statistical methodologies evolving just as the need is becoming so great. Many of these are coming from the scientific communities in Santa Fe and Los Alamos, New Mexico, where scientists have been working for years to take the advanced models developed for applications of deep physics and apply them to the practical problems faced by organizations like insurance companies. These new tools are designed specifically to analyze non-linear complex systems. Their underlying assumptions include the certainty of multiple correlations and cascading events, features that are the hallmark of complex systems. They are, in fact, the tools of complexity science applied to actuarial science.

Applied Complexity Science

Complexity science, the study of complex systems, has been developing as a discipline for several decades. The mathematical origins of its primary analytical tool, simulation technology, go back at least half a century, to the work of John Louis von Neumann and Stanislaw Ulam in the 1940s and 1950s. Von Neumann was a brilliant mathematician who made seminal contributions to the field of computing, including applications of mathematics to computing, and the application of computing to such disciplines as mathematical physics and economics. Ulam is known as the mathematician who solved the problem of how to initiate fusion in the hydrogen bomb and devised the "Monte-Carlo method" widely used in solving mathematical problems using statistical sampling.

Complex Systems

The global insurance market is what scientists call a "complex system." Complex systems have been extensively studied and exhaustively defined over the past several decades. Complex systems-things like atoms, molecules, economies and insurance markets-consist of a large number of individual agents that behave in accordance with certain basic motivations, rendered as "rules." Agents change their behavior on the basis of information they receive about what the other agents in the system are doing in order to continue to adhere to their motivating rules and adapt to the changing environment.

Artificial Intelligence

The theoretical work of the 1950s mathematicians was picked up by others and spawned a broad spectrum of new analytical technology. A technology that showed brief promise early on was labeled "artificial intelligence." Researchers thought they could create artificial intelligence by capturing all of the rules for doing something. Although the promise of artificial intelligence as originally envisioned was never realized, the research led to two important and robust modern technologies: expert systems and machine learning.

Expert Systems

Expert systems are the rules-driven technological descendants of artificial intelligence. In the 1960s, linguistics contributed the concept of fuzzy logic, a superset of Boolean logic designed to handle values that are neither completely true nor completely false. This led to the development of fuzzy expert systems, which came into widespread use in the early 1990s primarily as control and data analysis systems using a collection of fuzzy membership functions and rules. Today, these survive as knowledge-based expert systems, application programs that make decisions or solve problems in a particular field by using stored knowledge and analytical rules defined by experts in the field. Although such systems can become extremely complex and encyclopedic in the scope of the information to which they provide access, they do not help us to analyze complex systems because they do not address the inter-relationships between the various pieces of information.

Machine Learning

Machine learning—the ability of a computer system to autonomously acquire and integrate knowledge—is fundamental to the study and analysis of complex systems. Machine learning is the capacity to learn from experience, analytical observation and other means, and therefore to continuously self-improve. It makes possible adaptive agent-based simulation technology, which is the best tool available today for the study of complex systems. Simulation technology is not new: for decades, scientists have used it to study many aspects of the natural world that seemed closed to traditional scientific methods.

Particle Mathematics

Insights coming from discoveries in modern physics are breathtaking, and they are beginning to influence actuarial science. The mathematics physicists have used to penetrate the dual (wave-particle) nature of subatomic physical reality is now proving to be applicable to practical problems of modern finance. Mathematical tools used by physicists to probe the equivalence of matter and energy, the merging of space and time, the interconnectedness of disparate parts of the cosmos, are the same ones used by actuaries to trace the connections between contingencies in the world and financial consequences in society.

In particular, scientists who have become interested in complex adaptive systems have begun to uncover patterns from nature which are duplicated in the world of finance. From data mining and visualization through neural networks and genetic algorithms, patterns of self-organization, emergence, fractal structures and dynamic landscapes have proven to be vital in interpreting patterns that influence the outcome of risk management activities.

Chaos Theory

Chaos theory had discovered that simple natural phenomena like clouds, coastlines and flowers did not conform to the mathematical structures scientists had created by which to analyze them.

By allowing natural phenomena to define themselves using advanced computer technology, scientists were able to create new models of nature which conformed more closely to actual nature than the nature we imposed on nature.

From Laboratory Development to Desktop Application

Until recently, simulation technology was largely confined to the supercomputing facilities of facilities like the Los Alamos National Laboratory because the huge amounts of data required to build a useful simulation could only be processed on large supercomputing platforms. Today, the processing power to create a realistic simulation of a restricted environment can be packaged in a personal computer or even a laptop. This revolution in the availability of computing power has spurred the growth of a burgeoning industry in applied complexity science.

Complexity scientists in Los Alamos and Santa Fe have harnessed the power of computer and analytical methodologies used by leading physicists to apply to practical problems of business. One area of recent business stress which is especially amenable to these technologies is risk analysis. The dynamic, nonlinear, agent based methodologies used by these scientists in finding patterns in complex adaptive systems are ideally suited to a world of extreme events and high correlation.

Complexity science and its sister area of chaos theory have brought the power of the computer and advanced mathematics to the world of risk. Because they allow data to form its own categories (using mechanisms like neural networks) and because they allow analytics to form around the emerging data structures, they avoid the rigidity of traditional linear models. In the case of extreme events and unusual correlations, this is a way to better address the real world of risk rather than the one which we assume exists.

Analytical Tools for Complex Systems

Complex systems are best studied through the technology of adaptive agent-based simulation, a technology expressly designed to analyze and model them. Adaptive agent-based simulation technology depends on a new breed of risk assessment and management tools developed from the practical application of complex systems simulation technology. Originally designed to detect patterns that exist within complex adaptive systems like the human body or the atom, these deep methodologies were found to apply to the very practical problems with which businesses and other institutions were struggling.

Adaptive Agent-Based Simulation

Adaptive agent-based simulation builds a modeling system that permits the modeler to keep track of and modify the behavior of each individual in a synthetic population. It uses all of the modern mathematical techniques described above, but depends especially on machine learning to develop reliable simulations. Simulators permit modeling in real time and demonstrate interactions and correlations between multiple events in real time.

Inside the Black Box

One of the barriers to widespread promulgation of adaptive agent-based simulation technology is the "black box" mystique that sometimes surrounds it. While it is true that the science and mathematics behind it is very complex, some of the families of algorithms and analytical techniques that were developed to implement machine learning are familiar to many, if not all, actuaries, though they may be applied in very different ways. In this section of the paper, we examine some of the nuts and bolts that go into building the black box.

Data Mining

The proliferation of data has led to the rapid emergence of new technologies and disciplines focused on how to store, retrieve and use it. While data has been accumulating in mainframe systems for 50 years, it was barely a decade ago that the first client-server terabyte data storage banks were deployed. Moving these data storage banks, now commonly called data warehouses, to the client-server environment greatly increased flexibility in querying the data. Data warehouses were the strategic prerequisites to data marts, tactical data repositories designed for ease of access and usability for a particular purpose and to meet an immediate need.

The discipline of knowledge management emerged as businesses struggled to organize these huge volumes of data into information and hence, it was hoped, into knowledge. This led to the technology of data mining, the process of extracting valid, useful, previously unknown and ultimately comprehensible information from data warehouses. Data mining includes the classification, clustering and segmenting of the data, as well as the detection of rules of association, sequential patterns and deviations.

Data Fusion

Until recently, a prerequisite to successful data mining was data normalization: that is, the data to be mined would have to be housed in a single database, or, if distributed, in databases of parallel structure. One of the advances that has made adaptive agent-based simulations easily accessible for business applications is the evolution of software agents that are able to mine disparate databases for patterns and correlate those patterns outside the structure of an individual database.

This technique, called data fusion, means that companies can correlate information from legacy mainframe databases, client-server CRM systems and purchased census or marketing data without porting any of it to a different system or structure. The mathematical techniques used by individual agents to mine the data are not unfamiliar—they include, for example, cluster correlations, logistic regression and partial least squares. The ability to correlate non-normalized data, however, is revolutionary.

Building Non-linear Models

Though data mining is an advanced technology, by and large it remains mired in the linear mind-set. Non-linear correlations defy traditional linear actuarial mathematics. Since complex systems are, as we have said, non-linear in nature, data mining can only provide limited insight into such systems. More often than not, in non-linear systems causes lead to effects that appear entirely counterintuitive. For example, a group of researchers at Los Alamos built a transportation system simulation that would display both the traffic congestion and the air pollution effects of creating incentives for people to use public transportation. The simulation revealed that air pollution will increase when car traffic decreases in certain situations. The finding was validated—it derived from the traffic pattern shifting to many shorter trips, so that the pollution-controlling catalytic converters in most cars never warmed up enough to affect emissions—but it was certainly a surprise to traffic planners!

The adaptive agent-based simulation technology was developed specifically to analyze non-linear systems. It incorporates a number of new mathematical tools expressly created for dealing with complex systems. It also incorporates what researchers have learned about human perception in the past half-century to order the information in ways that are more accessible to human perception. Both the analysis and the visualization tools are necessary to reduce these unimaginable volumes of data to information that is meaningful in human terms and, ultimately, to knowledge. We will briefly touch on some of the more well-known or significant of these tools.

Cellular Automata

Cellular automata are commonly mentioned in connection with complexity science and have come to represent complexity in the popular culture. The most basic cellular automata play a simple "game" in which you have a row of agents, and each agent can be black or white. Agents determine whether they will be black or white by looking at what their neighbors are doing. So if an agent has a black neighbor both to the right to the left, there might be a rule that says: "If my neighbors are both black, I'm going to be white." Of course, neighboring agents also have rules and are following them, so that the environment continues to evolve.

Cellular automata are fascinating because very complicated behaviors, a kaleidoscope of patterns, can form from very few, very simple rules. Mathematically, cellular automata illustrate the ability of local parallel update rules to generate spatial structure from disordered initial states. The mathematics of cellular automata can also be traced back to the work of von Neumann and Ulam in the 1940s. Today, the game LIFE, invented by Cambridge mathematician John Conway, is the most popular illustration of cellular automata. It presents a simple two-dimensional analog of basic processes in living systems. The game traces changes through time in the patterns formed by sets of adaptive cells arranged in a two-dimensional grid. Any cell in the grid may be either "alive" or "dead." The state of each cell changes from one generation to the next depending on the state of its immediate neighbors, according to a simple set of four rules.

The behavior of LIFE illustrates how cellular automata reproduce the tendency toward order of living systems. Starting from an arbitrary initial configuration, order (pattern) usually emerges fairly quickly. Ultimately, most configurations either disappear entirely or break up into isolated static or cyclical patterns. This and other games that illustrate different mathematical concepts in cellular automata can be found at: http://psoup.math.wisc.edu/sink.html. These concepts are mathematically fascinating, but so far have given rise to very few practical applications, and these mostly in the area of modeling natural systems. Complexity scientists who deal with business simulations do not generally consider cellular automata particularly useful.

Neural Networks and Genetic Algorithms

Neural networks and genetic algorithms are critical building blocks for non-linear models. Loosely based on the human brain's physical pattern of learning, neural networks receive data as input and produce output in the form of behavior. Genetic algorithms are a class of heuristic optimization methods and computational models of adaptation and evolution based on Darwinian natural selection. Just as neural networks mimic the activity of the brain, genetic algorithms are based on DNA and genetics. The agents of a simulation behave in reaction to the inputs. Driven by their genetic algorithms, they survive, die or dominate. Those who die are eliminated from the simulation, those who survive are retained and mated, and those who dominate – the superheroes – are emulated.

Data Visualization

Data visualization is critical to building useful simulations. It allows scientists and non-scientists alike to explore trends within a body of data by visually orienting themselves to the patterns in the data. Because it can help translate data patterns into insights, data visualization is a highly effective decision-making tool. It can represent scenario results in graphic format or in the form of pro-forma financial statements or in any other format that makes sense in the context of the industry being modeled. A common use is to aggregate inconceivable amounts of data into patterns that make sense to the eye.

Simulation for Insurance Companies

Complexity science uses the emerging power of computer technology and mathematics to allow scientists to take data from society and nature and let the data create its own categories. This allows scientists to approach reality before constraining it to the categories we have created for it. Data mining techniques of high sensitivity, algorithms incorporating elements of neural networks, genetic algorithms, cellular automata, partial least squares, logistic regression and advanced data visualization processes are all ready for deployment in the insurance industry. They have been tested and found to be robust and efficient in the most trying of circumstances.

The adaptive agent-based simulation tools developed for actuaries by complexity scientists complement traditional DFA and ERM modeling approaches. They help provide a global, integrated look at a company's risk profile. They can be used strategically to test different growth and hedge strategies over a long time horizon given various scenarios about extreme events and cross correlations.

They can be used to support pricing and capital allocation decisions as well as reinsurance and diversification programs. Risk-Return analysis, line of business allocations, demographic structure and product design can all benefit from such powerful perspective. The power of decades of effort in the most powerful research institutions in the world is brought to the desktop of the contemporary insurance executive.

Insurance Applications of Simulation Technology

Every type of insurance coverage has unique issues that have arisen in recent years. Adaptive agent-based simulations have been developed to meet insurers' need for models that incorporate the dynamics of insurance instead of trying to fit data into historic, static techniques. Operating without the limitations of linear thinking, adaptive agent-based simulation can model correlative and cascading effects on unprecedented scales. Rather than presume outcomes or statistical limits, adaptive agent-based simulations enable the key influencing factors (agents) to shape the outcome. The subsequent model is a more responsive, more accurate reflection of today's complex risk dynamics.

Simulation for Property and Casualty Insurers

A simulation of the impact of catastrophic events on the insurance industry has been developed using the most advanced tools available to complexity scientists. Assuratech's *InsuranceWorld©* simulates how various insurers, reinsurers and capital markets in general will fare under different catastrophic scenarios. It can be used by an individual company to strategize how to position itself relative to competition, or by regulators or others interested in some of the larger issues involved in extreme event coverage.

The *InsuranceWorld©* simulation is populated by five primary insurers and five reinsurers who operate in four or more markets within a described economy and are impacted by three or more extreme events (natural or man-made catastrophes). Using selected parameters and strategies for the companies and certain scenarios for the economy and nature, it produces financial statements for companies and global outputs. For example, a company might use the model to develop hedging strategies by comparing projected financials with and without the strategies.

A specific terrorist component can be added to the simulation to show various worldwide assets vulnerable to attack, various scenarios for insurance coverage, various events which could affect those assets and coverages and various ways of adjusting. In the area of policy and planning, such a module could help model how government and industry can share layers and slices of terrorism risk. It could also provide advanced maximum insured loss scenarios.

By implementing Assuratech's agent-based, nonlinear modeling, insurers can test for the impact of extreme events and high correlations in ways never before possible. As a result, senior executives can protect their companies and earnings flow from disruptions that have never before been hedgeable—a significant market advantage for the first user of this technology.

Simulation for Workers' Compensation Insurers

Following the pattern of the *InsuranceWorld©* simulator, similar simulators are being developed to model other insurance markets. A market in dire need of this technology is the workers' compensation market. A convergence of events, some natural, some human-generated, has brought the workers' compensation system to a crisis. Workers' compensation rates determine a major cost of doing business in a state, and as a result are always under scrutiny by State Insurance Departments and other government agencies. Self-insurance is an option often chosen by businesses, which further complicates the environment in which workers' compensation operates.

Fraud has always been a problem, as worker injuries are not always easy to determine objectively. Because of the time lag between insurance coverage and the payout of associated claims, there is uncertainty about the true liabilities at any point in time. As if these traditional issues related to workers compensation were not enough, we now impose the complexity of terrorism, globalism, and the information revolution. Workers' compensation professionals are looking for new perspectives and tools by which to help manage the dynamic risk landscape that they now face.

A market simulator for the workers' compensation industry would provide senior management strategic decision-support information to enable them to anticipate the repositioning of the company forward of major swings in the market. The tool can examine external market influences in a comprehensive manner and reliably estimate their effect on an insurer's market share. It demonstrates in the familiar terms of a financial statement the cascading effect of dramatic swings in market share on such factors as pricing, investment performance, solvency and overall company fiscal health. It can also demonstrate the effects of extreme events such as natural disasters and terrorism. This tool can easily accommodate the additional uncertainty that stems from longer-tailed liabilities and the impact of immature loss ratios on company financials.

The Emerging Paradigm

In recent years, actuarial periodicals have seen an explosion of articles incorporating elements of the new math. From multifractal modeling and non-linear perspective through neural networks and dynamic pricing, actuaries are beginning to utilize new areas of modern mathematics.

Given the underlying nature of actuarial phenomena, this is unavoidable. Differential equations that portray physical phenomena of the type modeled in actuarial applications are as complex as any faced by modern physicists. The uncertainty behind actuarial calculations trumps even that articulated by Heisenberg. The probabilities behind the financial impacts of actuarial events make those of particles hidden in Schrodinger's wave equation seem manageable.

By bringing the force of the advanced technology and mathematics of recent years into the actuarial paradigm, actuaries will see significant improvements in modeling capabilities. Problems felt to be intractable will open up to the new perspective. The challenges of the new global economy will again be matched against analytical tools capable of meeting them.

Conclusion

Adaptive agent-based simulation technology rests on the simple premise that the aggregate statistical behavior of market or population segments (the top-down view) is the result of the behavior of the individual agents that comprise that segment ***and their interactions.*** The tools developed by complexity scientists to model agent behavior in the insurance market represent a new paradigm in risk modeling. They address emerging issues and encompass extreme events, multiple correlations and other revolutionary changes in market environment that have recently plagued the financial community.

The mathematics behind the new tools is the same as that at the foundation of actuarial mathematics. Most applied mathematics involves finding functions which fit phenomena of interest, and making adjustments thereto as needed. Actuarial phenomena, like physical phenomena, are not well-behaved, and techniques which can allow data to form its own categories, and which can account for the underlying complexity of interrelationships, will ultimately carry the day.

The Authors

Lee Smith, BS, MA, FCAS, MAAA

Mr. Smith has over thirty years' experience in consulting, industry, and academia. The majority of his experience involves areas of financial services and insurance. He has served as Director of Planning at Aetna Insurance Company, Chief Actuary and Policy Director for the Michigan Insurance Bureau, Adjunct Professor of Actuarial Science at the University of Michigan, President of Midwest Casualty Actuaries, and National Director of Ernst & Young's Actuarial and Risk Management Practice. He can be reached at LeeSmithM@aol.com.

Lilli Segre Tossani, BA, MA

Lilli Segre-Tossani is a technical writer who has worked for fifteen years in the fields of information technology, financial and legal services, and organizational consulting. She can be reached at tossani@attglobal.net.

Overview of Enterprise Risk Management

The CAS Enterprise Risk Management Committee

Overview of Enterprise Risk Management

Casualty Actuarial Society

Enterprise Risk Management Committee

May 2003

Table of Contents

Appendices

Overview of Enterprise Risk Management

I. Executive Summary

This document is intended primarily to further the risk management education of candidates for membership in the Casualty Actuarial Society (CAS). Current members of the CAS as well as other risk management professionals should also find this material of interest.

In Chapter II, the evolution to and rationale for enterprise risk management (ERM) is explained. The "ERM movement" is driven by both internal (e.g., competitive advantage) and external (e.g., corporate governance) pressures – pressures that are both fundamental and enduring.

Chapter III defines ERM for CAS purposes, and lays out its conceptual framework. The definition makes clear that ERM is a value-creating discipline. The framework describes both the categories of risk and the types of risk management processes covered by ERM. ERM is seen to extend well beyond the hazard risks with which casualty actuaries are particularly familiar, and well beyond the quantification of risks with which they are particularly skilled – but it is clear that the casualty actuarial skill set is extremely well-suited to the practice of ERM. ERM also extends well beyond the insurance industry, which presents a distinct opportunity for casualty actuaries to continue to expand their career horizons and take leadership roles in these varied industries.

The vocabulary of ERM is established in Chapter IV, which also describes the measures, models and tools supporting the discipline. The close linkage between ERM and corporate performance management is made clear in this discussion. Dynamic Financial Analysis (DFA) is introduced, along with alternative approaches to capture hazard and financial risks, and their roles within an ERM context is explained. Models that treat operational and strategic risks are also discussed. Applications of these measures, models and tools to support management decision-making are outlined at the conclusion of this chapter.

With the conceptual and technical foundations of ERM thus established, Chapters V and VI turn to the actual practice of ERM. Chapter V presents relevant case studies from various industries, and Chapter VI offers some practical considerations in implementing ERM.

For the reader interested in pursuing additional sources of learning on the subject, a bibliography of existing literature on ERM and its key components is included in Appendix C. (A continually updated, annotated and topically-organized road map through the literature can be found on the CAS Web site at http://www.casact.org/research/erm/.)

Enterprise risk management is a "big idea". Among other things, ERM can be viewed as the broad conceptual framework that unifies the many varied parts of the actuarial discipline. ERM provides a logical structure to link these subject areas together in a compelling way to form an integrated whole. In so doing, ERM addresses critical business issues such as growth, return, consistency and value creation. It expresses risk not just as threat, but as opportunity – the fundamental reason that business is conducted in a free enterprise system. Through ERM, the clear linkage between business fundamentals and actuarial theory and practice should engage students and professionals from various backgrounds in the study of actuarial science – a logical career strategy in a global business environment that has embraced ERM as a modern management discipline.

Overview of Enterprise Risk Management

II. The ERM Evolution

Organizations have long practiced various parts of what has come to be called enterprise risk management. Identifying and prioritizing risks, either with foresight or following a disaster, has long been a standard management activity. Treating risks by transfer, through insurance or other financial products, has also been common practice, as has contingency planning and crisis management.

What has changed, beginning very near the close of the last century, is treating the vast variety of risks in a holistic manner, and elevating risk management to a senior management responsibility. Although practices have not progressed uniformly through different industries and different organizations, the general evolution toward ERM can be characterized by a number of driving forces. We discuss these characteristic forces below.

More – and More Complicated – Risks

First of all, there is a greater recognition of the variety, the increasing number, and the interaction of risks facing organizations. Hazard risks such as the threat of fire to a production facility or liability from goods and services sold have been actively managed for a long time. Financial risks have grown in importance over the past number of years. New risks emerge with the changing business environment (e.g., foreign exchange risk with growing globalization). More recently, the awareness of operational and strategic risks has increased due to a succession of high-profile cases of organizations crippled or destroyed by failure of control mechanisms (e.g., Barings Bank, Enron) or by insufficient understanding of the dynamics of their business (e.g., Long Term Capital Management, General American Insurance Company). The advance of technology, the accelerating pace of business, globalization, increasing financial sophistication and the uncertainty of irrational terrorist activity all contribute to the growing number and complexity of risks. It is reasonable to expect that this trend will continue.

Organizations have come to recognize the importance of managing *all* risks and their interactions, not just the familiar risks, or the ones that are easy to quantify. Even seemingly insignificant risks on their own have the potential, as they interact with other events and conditions, to cause great damage.

External Pressures

Motivated in part by the well-publicized catastrophic failures of corporate risk management cited above, regulators, rating agencies, stock exchanges, institutional investors and corporate governance oversight bodies have come to insist that company senior management take greater responsibility for managing risks on an enterprise-wide scale. These efforts span virtually every country in the civilized world. A sampling of these requirements and guidelines has been compiled in Appendix A.

In addition to these codified pressures, publicly traded companies are well aware of the increasingly vocal desire of their shareholders for stable and predictable earnings, which is one of the key objectives of ERM for many organizations.

Portfolio Point of View

Another characteristic force is the increasing tendency toward an integrated or holistic view of risks. Developments in finance (i.e., Modern Portfolio Theory) provide a framework for thinking about the collective risk of a group of financial instruments and an individual security's contribution to that collective risk. With ERM, these concepts have been generalized beyond financial risks to include risks of all kinds, i.e., beyond a portfolio of equity investments to the entire collection of risks an organization faces. A number of principles follow from this thinking, including:

- Portfolio risk is not the simple sum of the individual risk elements.
- To understand portfolio risk, one must understand the risks of the individual elements plus their interactions.
- The portfolio risk, or risk to the entire organization, is relevant to the key risk decisions facing that organization.

The implications of these principles are having a significant impact on the practice of ERM. There is growing recognition that risks must be managed with the total organization in mind. To do otherwise (sometimes referred to as managing risk within "silos") is inefficient at best, and can be counter-productive. For example, certain risks can represent "natural hedges" against each other (if they are sufficiently negatively correlated). A classic case is that of an insurer selling both life insurance and annuity business to similarly situated customers and thereby naturally hedging away its mortality risk. To separately hedge mortality risk on these products (e.g., through reinsurance) would be cost inefficient and entirely unnecessary. Another example is that of a global conglomerate with one of its divisions long in a certain foreign currency and another short in the same currency. Separate currency hedges, while seemingly advisable from the point of view of the individual division heads, are unreasonable for the enterprise as a whole.

A holistic approach helps give organizations a true perspective on the magnitude and importance of different risks.

Quantification

A fourth characteristic force, closely tied to the third, is the growing tendency to quantify risks. Advances in technology and expertise have made quantification easier, even for the infrequent, unpredictable risks that historically have been difficult to quantify. Following a series of natural disasters, most notably hurricane Andrew in 1992, the practice of catastrophe modeling arose and is now a standard practice in insurance companies. This combination of meteorological (in the case of hurricane modeling), structural engineering, insurance and technological expertise leading to probabilistic models is a huge advancement over previous quantification attempts. By the end of the twentieth century, insurance and reinsurance companies routinely measured their exposure to hurricanes, earthquakes and other natural disasters with a greater degree of precision leading to a greater confidence in the ability to manage the exposure. More recently, such exposure-based quantification of exposure to losses has been extended to even less predictable, man-made disasters such as terrorist attacks.

The emergence of Value-at-Risk as a regulatory and management standard in the financial services industry has been aided by the speed and ease in measuring certain financial risks. Data is collected constantly allowing risk profiles to be adjusted as portfolios and market conditions change. This gives financial institutions and the regulatory bodies that oversee them a level of confidence in their ability to take actions to operate within established parameters.

Despite these advances, there will always remain risks that are not easily quantifiable. These include risks that are not well defined, unpredictable as to frequency, amount or location, risks subject to manipulation and human intervention, and newer risks. Man-made risks, operational and strategic risks are examples of these. Operational risk is a general category for a wide variety of risks, many of which are influenced by people and many of which do not have a long historical record. The tendency to quantify exposure to all these risks will certainly continue.

In the same way there has been a continuing effort to better quantify individual risks, there is a growing effort to quantify portfolio risk. This effort is much more difficult because in addition to individual risks, one must quantify or explain interactions between individual risk elements. This can be extremely complex and challenging. However, there often is not the need for a great deal of precision; even a directionally correct answer may be valuable. The attempt at quantification allows the organization to analyze "what if" scenarios. They are able to estimate the magnitude of risk or degree of dependency with other risks sufficiently to make informed decisions. Further, simply going through the quantification process gives people a better qualitative perspective of the risk. They may gain insight as to the likelihood or severity of the risk or to ways to prevent or mitigate the exposure.

Boundaryless Benchmarking

A fifth characteristic force pertains to scope. Common ERM practices and tools are shared across a wide variety of organizations and across the globe. The process, tools, and procedures laid out in this overview are not limited to the insurance or even financial service industries but rather are common to many organizations. Information sharing has been aided by technology but perhaps more importantly, because these practices are transferable across organizations. Organizations have become quite willing to share practices and efficiency gains with others with whom they are not direct competitors.

An example of a phenomenon common to many organizations and having risk management implications is real options. Many organizations face operating and strategic situations where events are uncertain, players make initial investments to get in the game and then have the opportunity to make successive investments contingent on future events. The drug approval process in the pharmaceutical industry is an example where organizations face options-like decisions (see Chapter V). Option pricing techniques provide organizations with a means of better thinking about and managing these risks.

Different industries and organizations will continue to develop and employ variations of ERM. Different risks will be more or less important to organizations and risk management practices will differ in particular ways that best suit the organization, but there will be general concepts and broad general practices and techniques that are recognized and employed by organizations throughout the world.

Risk as Opportunity

A sixth characteristic force pertains to the outlook organizations have toward risk. In the past, organizations tended to take a defensive posture towards risks, viewing them as situations to be minimized or avoided. Increasingly, organizations have come to recognize the opportunistic side, the value-creating potential of risk. While avoidance or minimization remain legitimate strategies for dealing with certain risks, by certain organizations at certain times, there is also the opportunity to swap, keep, and actively pursue other risks because of confidence in the organization's special ability to exploit those risks.

There are a number of reasons for this shift in attitude. Over time and with practice, organizations have become more familiar with and more capable of managing the risks they face. They develop expertise in managing those risks both because of familiarity and confidence in the organization's abilities. As a result, they may keep their own exposure and seek out opportunities to assume other organization's risks. Over time, better information about risk has become available. This has led to new markets for trading risks and more information about the cost of risks. This has allowed organizations to better evaluate risk and return trade-offs and see that the costs of transfer sometime outweigh the benefits. In addition, the existence of risk-trading markets contributes to a greater degree of confidence. Organizations can adopt a more aggressive stance if they know they can switch to a defensive stance quickly, if needed.

In some cases organizations seek out risks to increase diversification, realizing that the addition of some risks may have a minimal impact on overall risk, or in the case of hedges, may decrease enterprise risks. In essence, there is a realization that risk is not completely avoidable and, in fact, informed risk-taking is a means to competitive advantage.

Summary

It is reasonable to expect that the forces cited above will continue. Accordingly, risk management practices will become more and more sophisticated. As capabilities continue to improve, organizations will increasingly adopt ERM because they can.

* * * * *

Overview of Enterprise Risk Management

Note: For additional thoughts on the subject of this chapter, see Lisa K. Meulbroek, "Integrated Risk Management for the Firm: A Senior Manager's Guide", *Harvard Business School's Division of Research Working Papers 2001-2002*, http://www.hbs.edu/research/facpubs/workingpapers/papers2/0102/02-046.pdf.

Overview of Enterprise Risk Management

III. ERM Definition and Conceptual Framework

Definition

Several texts and periodicals have introduced or discussed concepts such as "strategic risk management", "integrated risk management" and "holistic risk management". These concepts are similar to, even synonymous with, ERM in that they all emphasize a comprehensive view of risk and risk management, a movement away from the "silo" approach of managing different risks within an organization separately and distinctly, and the view that risk management can be a value-creating, in addition to a risk-mitigating, process.

The CAS Committee on Enterprise Risk Management has adopted the following definition of ERM:

> **"ERM is the discipline by which an organization in any industry assesses, controls, exploits, finances, and monitors risks from all sources for the purpose of increasing the organization's short- and long-term value to its stakeholders."**

Several parts of this definition merit individual attention. First, ERM is a discipline. This is meant to convey that ERM is an orderly or prescribed conduct or pattern of behavior for an enterprise, that it has the full support and commitment of the management of the enterprise, that it influences corporate decision-making, and that it ultimately becomes part of the culture of that enterprise. Second, ERM, even as it is defined for CAS purposes, applies to all industries, not just the property/casualty insurance industry with which casualty actuaries are intimately familiar. Third, the specific mention of exploiting risk as a part of the risk management process (along with the stated objective of increasing short- and long-term value) demonstrates that the intention of ERM is to be value creating as well as risk mitigating. Fourth, all sources of risk are considered, not only the hazard risk with which casualty actuaries are particularly familiar, or those traditionally managed within an enterprise (such as financial risk). Lastly, ERM considers all stakeholders of the enterprise, which include shareholders and debtholders, management and officers, employees, customers, and the community within which the enterprise resides.

Implicit in this definition is the recognition of ERM as a strategic decision support framework for management. It improves decision-making at all levels of the organization.

Conceptual Framework

A useful way to conceptualize ERM is along two dimensions: one spanning the *types* of risk included, and the other spanning the various risk management *process* steps, as below:

Overview of Enterprise Risk Management

ERM Framework				
	Types of Risk			
Process Steps	Hazard	Financial	Operational	Strategic
Establish Context				
Identify Risks				
Analyze/Quantify Risks				
Integrate Risks				
Assess/Prioritize Risks				
Treat/Exploit Risks				
Monitor & Review				

In discussing these risk types and process steps, we will consider an enterprise, the Coldhard Steel Company ("Coldhard Steel"), which manufactures steel products, such as roller and ball bearings, used in other industrial machinery. Coldhard Steel operates in the "rust belt" of the midwestern U.S., is family-owned, and has a unionized labor force.

Types of Risk

Coldhard Steel is exposed to a number of ***hazard risks***. First-party hazard risks include the possibility of fire or tornadoes damaging its plant and equipment, and the resulting loss of revenue (i.e., business interruption). Second-party hazard risks include injury or illness to its employees, including work-related injuries that would result in workers compensation claims. Given Coldhard Steel's use of heavy machinery, as well as the benefit provisions in its principal state of operation, Coldhard Steel's workers compensation exposure is substantial. Third-party hazard risk would include the possibility of slips and falls of visitors on its premises, products recall and/or products liability from defective products produced by Coldhard Steel.

Since Coldhard Steel has significant sales in Latin America and Europe, it is exposed to foreign exchange risk, one of many ***financial risks***. Coldhard Steel is tangentially exposed to additional foreign exchange risk in that even though it buys its steel from U.S. manufacturers, these prices are influenced by imported steel. Other financial risks for Coldhard Steel to consider are commodity risk (due to possible changes in prices in the raw materials it and its suppliers use in production) and credit risk (due to its significant accounts receivables asset).

Since many employees are in the local machinists union, labor relations represents a significant ***operational risk*** for Coldhard Steel. Also, since the company is privately held, succession planning is critical for the time when the current owner either sells the company or passes down control to heirs. Coldhard Steel spends considerable time assessing the efficiency and reliability of its machines and processes.

Strategic risks for Coldhard Steel include fluctuations in the demand and the market price for its finished products (and substitute products), competition from suppliers of other steel products, regulatory/political issues associated with the steel industry, and technological advances in its customers' machines that could potentially render Coldhard Steel's current products obsolete.

In general, enterprises (like and unlike Coldhard Steel) are exposed to risks that can be categorized into the following four types:

- *Hazard Risks* include risks from:
  - fire and other property damage,
  - windstorm and other natural perils,
  - theft and other crime, personal injury,
  - business interruption,
  - disease and disability (including work-related injuries and diseases), and
  - liability claims.

- *Financial Risks* include risks from:
  - price (e.g. asset value, interest rate, foreign exchange, commodity),
  - liquidity (e.g. cash flow, call risk, opportunity cost),
  - credit (e.g. default, downgrade),
  - inflation/purchasing power, and
  - hedging/basis risk.

- *Operational Risks* include risks from:
  - business operations (e.g., human resources, product development, capacity, efficiency, product/service failure, channel management, supply chain management, business cyclicality),
  - empowerment (e.g., leadership, change readiness),
  - information technology (e.g., relevance, availability), and
  - information/business reporting (e.g., budgeting and planning, accounting information, pension fund, investment evaluation, taxation).

- *Strategic Risks* include risks from:
  - reputational damage (e.g., trademark/brand erosion, fraud, unfavorable publicity)
  - competition,
  - customer wants,
  - demographic and social/cultural trends,
  - technological innovation,
  - capital availability, and
  - regulatory and political trends.

The precise slotting of individual risk factors under each of these four categories is less important than the recognition that ERM covers all categories and all material risk factors that can influence the organization's value.

Overview of Enterprise Risk Management

Process Steps

The following steps of the risk management process, which are based on those originally detailed in the Australian/New Zealand Standard in Risk Management (AS/NZS 4360), describe seven iterative elements.

- *Establish Context* – This step includes External, Internal and Risk Management Contexts.

  - The External Context starts with a definition of the relationship of the enterprise with its environment, including identification of the enterprise's strengths, weaknesses, opportunities, and threats ("SWOT analysis"). This context-setting also identifies the various stakeholders (shareholders, employees, customers, community), as well as the communication policies with these stakeholders.

  - The Internal Context starts with an understanding of the overall objectives of the enterprise, its strategies to achieve those objectives and its key performance indicators. It also includes the organization's oversight and governance structure.

  - The Risk Management Context identifies the risk categories of relevance to the enterprise and the degree of coordination throughout the organization, including the adoption of common risk metrics.

  Returning to our example, Coldhard Steel has formed a Risk Management Committee that is headed by its chief financial officer, with representatives from loss control/safety, quality control, human resources, marketing, and finance. In consideration of the makeup of its labor force, a representative from the labor union is invited periodically to meetings. In terms of establishing common criteria for assessing all risks, Coldhard Steel adopted a Value at Risk approach, with an annual timeframe.

- *Identify Risks* – This step involves documenting the conditions and events (including "extreme events") that represent material threats to the enterprise's achievement of its objectives or represent areas to exploit for competitive advantage.

  In our example, Coldhard Steel has used a variety of methods (e.g., surveys, internal workshops, brainstorming sessions and internal auditing) to identify the significant hazard, financial, operational and strategic risks described in the previous section.

- *Analyze/Quantify Risks* – This step involves calibrating and, wherever possible, creating probability distributions of outcomes for each material risk. This step provides necessary input for subsequent steps, such as integrating and prioritizing risks. Analysis techniques range along a spectrum from qualitative to quantitative, with sensitivity analysis, scenario analysis, and/or simulation analysis applied where appropriate.

  As indicated previously, workers compensation represents a significant hazard risk for Coldhard Steel. However, it has a number of years of claims and exposure data, and, based on quantitatively extrapolating cost trends into the future, Coldhard Steel's consulting actuaries are able to determine reasonable expectations of costs and variability of these costs into the near future.

  Coldhard Steel regularly monitors its account sales and accounts receivables, including performing credit analysis on its largest customers before extending additional credit. Although all sales are transacted in U.S. dollars, orders from Mexico generate 10 percent of all sales, and Coldhard Steel's financial analysts have considered hedging against devaluations in the Mexican peso.

  Coldhard Steel's labor contract expires in three years, and although relations with the employees and union are considered good, senior management has asked its human resources to construct "best case", "expected" and "worst case" estimates of salary and benefit increases anticipated to be requested by labor. As part of the worst case scenario, management has asked its finance department to estimate the impacts of a prolonged labor dispute and its effects on revenue, expenses and inventories.

  Coldhard Steel buys its steel from U.S. manufacturers, even though some of its competitors are taking advantage of cheaper foreign steel. Coldhard Steel is actively monitoring political discussions to gauge the likelihood that additional tariffs will be imposed on foreign steel in the near future. Coldhard Steel also monitors price levels for its finished products in relationship to the cost of its raw materials, products of its competitors, and substitute products.

- *Integrate Risks* – This step involves aggregating all risk distributions, reflecting correlations and portfolio effects, and expressing the results in terms of the impact on the enterprise's key performance indicators (i.e., the "aggregate risk profile").

  Coldhard Steel's Risk Management Committee and external consultants have begun to develop a structural simulation model to integrate all risks. The various components of the model are supported by a common stochastic economic scenario generator.

- *Assess/Prioritize Risks* – This step involves determining the contribution of each risk to the aggregate risk profile, and prioritizing accordingly, so that decisions can be made as to the appropriate treatment.

Coldhard Steel has not yet quantified all risks into probability distributions, let alone integrated these risks into a complete aggregate risk profile. However, Coldhard Steel has developed judgmental assessments as to frequency and severity, and it has developed a "Risk Map", which plots all risks by these two components. Coldhard Steel has prioritized a number of risks including its workers compensation exposure (hazard), account bad debt/credit risk (financial), labor relation risk (operational), and product obsolescence risk (strategic).

- *Treat/Exploit Risks* – This step encompasses a number of different strategies, including decision as to avoid, retain (and finance), reduce, transfer, or exploit risk. For hazard risks, the prevalent transfer mechanism has been the insurance markets. Alternative risk transfer (ART) markets have developed from these with a goal of striking a balance between risk retention and risk transfer. With respect to financial risks, the capital markets have exploded over the last several decades to assist companies in dealing with commodity, interest rate, and foreign exchange risk. Until recently, companies had no mechanisms to transfer operational or strategic risks, and simply had to avoid or retain these risks.

  Coldhard Steel has historically insured its workers compensation exposure. However, given its comfort in assessing its loss experience, as well as increases in insurance rates, it is considering securing coverage with a large per occurrence deductible. With respect to financial risk, Coldhard Steel is instituting new standards regarding the extension of credit to its customers. In order to avoid potential labor disputes down the road, Coldhard Steel has decided to hold early discussions with union personnel regarding wages and benefits.

  Coldhard Steel believes that it is likely that additional tariffs will be imposed on foreign steel in the near future, so it is attempting to exploit this strategic risk by locking into fixed price agreements with its domestic suppliers.

- *Monitor & Review* – This step involves continual gauging of the risk environment and the performance of the risk management strategies. It also provides a context for considering risk that is scalable over a period of time (one quarter, one year, five years). The results of the ongoing reviews are fed back into the context-setting step and the cycle repeats.

  Coldhard Steel's newly formed Risk Management Committee met extensively toward the end of the previous year for planning purposes, and intends to meet monthly to monitor progress on goals established.

* * * * *

Note: The ERM Framework in this chapter was originally developed in the *Final Report of the Advisory Committee on Enterprise Risk Management* (the predecessor committee to the Enterprise Risk Management Committee). This November 2001 report is available on the CAS Web site at http://www.casact.org/research/erm/report.htm.

IV. ERM Language, Measures, Models and Tools

As outlined in the preceding chapter, the first process step in the ERM framework is to establish the context (internal, external and risk management) within which the organization operates. Critical to establishing this context – and one of the worthy goals of ERM in its own right – is the creation of a common risk vernacular across all functional areas and relevant disciplines throughout the organization. This chapter summarizes the terminology in common usage among companies that practice ERM, forming a large part the emerging global "language of risk". In so doing, this chapter introduces and discusses the measures, models and tools that help organizations perform the balance of the ERM process steps.

Where appropriate, certain items are compared and contrasted; and where some items represent alternative approaches to a similar issue, relative strengths and weaknesses are discussed.

Overall Corporate Performance Measures

ERM clearly links risk management with the creation of organizational value and expresses risk in terms of impact on organizational objectives. An important aspect of ERM is therefore the strong linkage between measures of risk and measures of overall organizational performance. Thus, our discussion of ERM terminology begins with a description of key corporate performance measures. Our focus is on publicly traded corporations, and where industry-specific details are introduced, we use the financial services industry (and, more specifically, the insurance industry) for illustration.

In addition to establishing context, these performance measures have specific application in the identification of risks. Risk identification is the qualitative determination of risks that are material, i.e., that potentially can impact, for better or worse, the organization's achievement of its financial and/or strategic objectives. These objectives are usually expressed, of course, in terms of the overall corporate performance measures.

The measures defined below are fundamental to the evaluation of corporate performance. It is assumed that the reader is already familiar with the more basic accounting terms and concepts such as net income, net worth, etc.

- General Industry
  - Return on equity (ROE) — net income divided by net worth.
  - Operating earnings — net income from continuing operations, excluding realized investment gains
  - Earnings before interest, dividends, taxes, depreciation and amortization (EBITDA) — a form of cash flow measure, useful for evaluating the operating performance of companies with high levels of debt (when the debt service costs may overwhelm other measures such as net income).
  - Cash flow return on investments (CFROI) — EBITDA divided by tangible assets.

- Weighted average cost of capital (WACC) — the sum of the required market returns of each component of corporate capitalization, weighted by that component's share of the total capitalization.
- Economic value added (EVA) — a corporate performance measure that stresses the ability to achieve returns above the firm's cost of capital. It is often stated as net operating profits after tax less the product of required capital times the firm's weighted average cost of capital.

- Financial Services Industry
  - Return on risk-adjusted capital (RORAC) — a target ROE measure in which the denominator is adjusted depending on the risk associated with the instrument or project.
  - Risk-adjusted return on capital (RAROC) — a target ROE measure in which the numerator is reduced depending on the risk associated with the instrument or project.
  - Risk-adjusted return on risk-adjusted capital (RARORAC) — a combination of RAROC and RORAC in which both the numerator and denominator are adjusted (for different risks).

- Insurance Industry
  - Economic capital — market value of assets minus fair value of liabilities. Used in practice as a risk-adjusted capital measure; specifically, the amount of capital required to meet an explicit solvency constraint (e.g., a certain probability of ruin).
  - RAROC — expected net income divided by economic capital (thus, the more technically correct label is RORAC – see above – but in the insurance industry, RAROC is the term commonly used). RAROC is typically employed to evaluate the relative performance of business segments that have different levels of solvency risk; the different levels of solvency risk are reflected in the denominator. Evaluating financial performance under RAROC calls for comparison to a benchmark return; when the benchmark return is risk-adjusted (e.g., for volatility in net income), the result is similar to RARORAC (see above), though the term RAROC is still applied.
  - Embedded value — a measure of the value of business currently on the books of an insurance company; it comprises adjusted net worth (the market value of assets supporting the surplus) plus the present value of expected future profits on in-force business. (Embedded value differs from appraisal value in that the latter also includes the value of future new business.) The performance measure is often expressed in terms of growth (i.e., year-on-year increase) in embedded value.
  - Risk Based Capital (RBC) — a specific regulatory capital requirement promulgated by the National Association of Insurance Commissioners. It is a formula-derived minimum capital standard that sets the points at which a state insurance commissioner is authorized and expected to take regulatory action.

Overview of Enterprise Risk Management

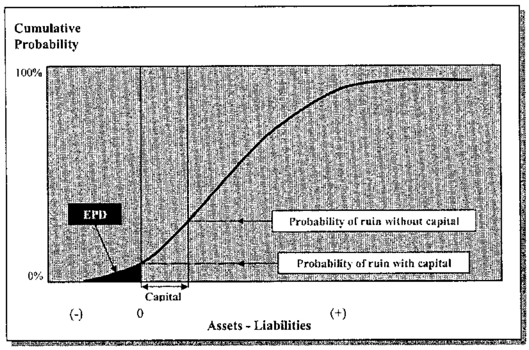
Risk Measures

In this section, reference is made to the term *"risk profile"* to represent the entire portfolio of risks that constitute the enterprise. Some companies represent this portfolio in terms of a cumulative probability distribution (e.g., of cumulative earnings) and use it as a base from which to determine the incremental impact (e.g., on required capital) of alternative strategies or decisions. It is in this sense that the term is used below.

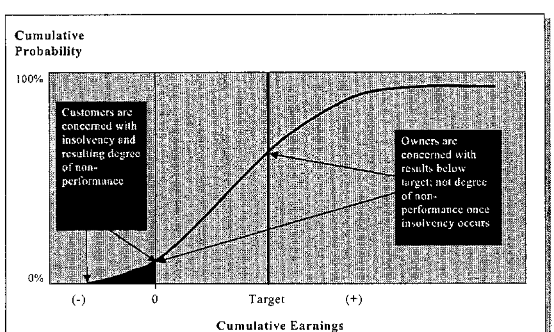
Most of the measures common in the practice of ERM can be placed in one of two categories: those measures related to the degree of the organization's *solvency*, and those related to the volatility of the organization's *performance* on a "going concern" basis. The measures in these two categories are used for distinctly different purposes and focus on distinctly different areas of the organization's risk profile. Following and complementing the narrative descriptions of these measures are illustrations and formulas where appropriate.

- *Solvency-related* measures (these measures concentrate on the adverse "tail" of the probability distribution – see "risk profile" above – and are relevant for determining economic capital requirements, i.e., they relate to the risks captured in the denominator of RARORAC; they are of particular concern to customers and their proxies, e.g., regulators and rating agencies):
  - Probability of ruin — the percentile of the probability distribution corresponding to the point at which capital is exhausted. Typically, a minimum acceptable probability of ruin is specified, and economic capital is derived therefrom.
  - Shortfall risk — the probability that a random variable falls below some specified threshold level. (Probability of ruin is a special case of shortfall risk in which the threshold level is the point at which capital is exhausted.)
  - Value at risk (VaR) — the maximum loss an organization can suffer, under normal market conditions, over a given period of time at a given probability level (technically, the inverse of the shortfall risk concept, in which the shortfall risk is specified, and the threshold level is derived therefrom). VaR is a common measure of risk in the banking sector, where it is typically calculated daily and used to monitor trading activity.
  - Expected policyholder deficit (EPD) or economic cost of ruin (ECOR) — an enhancement to the probability of ruin concept (and thus shortfall risk and VaR) in which the *severity* of ruin is also reflected. Technically, it is the expected value of the shortfall. (In an analogy to bond rating, it is comparable to considering the salvage value of a bond in addition to the probability of default.) For insurance companies, the more common term is EPD, and represents the expected shortage in the funds due to policyholders in the event of liquidation.
  - Tail Value at Risk (Tail VaR) or Tail Conditional Expectation (TCE) — an ECOR-like measure in the sense that both the probability and the cost of "tail events" are considered. It differs from ECOR in that it is the expected value, from first dollar, of all events beyond the tail threshold event, not just the shortfall amount.

- Tail events – unlikely but extreme events, usually from a skewed distribution. Rare outcomes, usually representing large monetary losses.



- *Performance-related* measures (these measures concentrate on the mid-region of the probability distribution –see "risk profile" above – i.e., the region near the mean, and are relevant for determination of the volatility around expected results, i.e., the numerator of RARORAC; they are of particular concern to owners and their proxies, e.g., stock analysts):
  - Variance — the average squared difference between a random variable and its mean.
  - Standard deviation — the square root of the variance.
  - Semi-variance and downside standard deviation — modifications of variance and standard deviation, respectively, in which only *unfavorable* deviations from a specified target level are considered in the calculation.
  - Below-target-risk (BTR) — the expected value of unfavorable deviations of a random variable from a specified target level (such as not meeting an earnings target).



Risk Measure	Formula
Standard Deviation	$\sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$ where n is the number of simulation iterations and *xbar* is the average value over all iterations. This is a commonly used measure of risk by academics and capital markets. It is interpreted as the extent to which the financial variable could deviate either above or below the expected value. Note that equal weight is given to deviations of the same magnitude regardless of whether the deviation is favorable or unfavorable. (There are different schools of thought on whether standard deviation in this context should measure total volatility or only the non-diversifiable volatility.)
Shortfall Risk *Value at Risk (VaR)*	$\frac{\sum_{i=1}^{n}[\textit{if } (x_i \le T) \quad \textit{then } 1, \textit{else } 0]}{n} * 100\%$ where T is the target value for the financial variable and n is the number of simulation iterations. This is an improvement over standard deviation because it reflects the fact that most people are risk averse, i.e., they are more concerned with unfavorable deviations rather than favorable deviations. It is interpreted as the probability that the financial variable falls below a specified target level. In VaR-type measures, the equation is reversed: the shortfall risk is specified first, and the corresponding value at risk (T) is solved for.
Downside Standard Deviation *Below Target Risk (BTR)*	$\sqrt{\frac{\sum_{i=1}^{n}(\min[0, (x_i - T)])^2}{n}}$ where T is the target value for the financial variable and n is the number of simulation iterations. This is a further improvement over the other metrics because it focuses not only on the probability of an unfavorable deviation in a financial variable (as with shortfall risk) but also the extent to which it is unfavorable. It is interpreted as the extent to which the financial variable could deviate below a specified target level. BTR is similar, but the argument is not squared, and there is no square root taken of the sum.

Risk Modeling

Risk modeling refers to the methods by which the risk and performance measures described above are determined. This chapter discusses the major classes of models used in the ERM process. It should be noted that these are general classes of models. The models used within any organization will typically be customized to accommodate the unique needs of, and the specific risks faced by, that organization. No two such models are exactly alike.

Most organizations will have at least a simple financial model of their operations that describes how various inputs (i.e., risk factors, conditions, strategies and tactics) will affect the key performance indicators (KPIs) used to manage the organization. For any given organization, these KPIs may be one or more of the overall corporate performance measures described earlier in this chapter (e.g., revenue growth, earnings growth, earnings per share, growth in surplus, growth in embedded value, customer satisfaction and/or brand image). For publicly traded companies, the KPIs are often explicitly or implicitly defined by the market (i.e., they are the measures focused upon by the organization's stock analysts). These models are often used in developing strategic and operational plans. For example, insurance companies typically make assumptions regarding future trends in claim costs by business segment (e.g., by line of business, by region), which are used to determine needed rate levels by segment. These rate level projections are then combined with assumptions on volume growth and other relevant inputs to derive a pro forma estimate of overall corporate earnings (or some other KPI). Often, business decisions (e.g., rate level, volume growth) are fine-tuned in order to produce the desired expected KPI result. Because these models explicitly capture the structure of the cause/effect relationships linking inputs to outcomes, they are termed *structural (or causal) financial models.*

These structural financial models are generally *deterministic* models because they describe *expected* outcomes from a given set of inputs without regard to the probabilities of outcomes above or below the expected values. These models can be transformed into *stochastic* (or *probabilistic*) models by treating certain inputs as variable. For example, expected future claim cost trend might be an input to a deterministic model of corporate earnings; recognizing that there is uncertainty in this trend, a probability distribution around the expected trend would be an input to a stochastic model. The model output, corporate earnings in this case, would then also be a probability distribution.

As outlined below, the two general classes of stochastic risk models are *statistical analytic models* and *structural simulation models*. "Statistical" vs. "structural" refers to the manner in which the relationships among random variables are represented in the model; "analytic" vs. "simulation" refers to the way in which the calculations are actually carried out. These four terms are defined separately below; the way they are combined is illustrated and contrasted in the table that follows the definitions.

- Analytic methods — models whose solutions can be determined "in closed form" by solving a set of equations. These methods usually require a restrictive set of assumptions and mathematically tractable assumed probability distributions. The principal advantage over simulation methods is ease and speed of calculation.

- Simulation methods (often called Monte Carlo methods) — models that require a large number of computer-generated "trials" to approximate an answer. These methods are relatively robust and flexible, can accommodate complex relationships (e.g., so-called "path dependent" relationships commonly found in options pricing), and depend less on simplifying assumptions and standardized probability distributions. The principal advantage over analytic methods is the ability to model virtually any real-world situation to a desired degree of precision.

- Statistical methods — models that are based on observed statistical qualities of (and among) random variables without regard to cause/effect relationships. The principal advantage over structural models is ease of model parameterization from available (often public) data.

- Mean/variance/covariance (MVC) methods — a special class of statistical methods that rely on only three parameters: mean, variance, and covariance matrix.

- Structural methods — models that are based on explicit cause/effect relationships, not simply statistical relationships such as correlations. The cause/effect linkages are typically derived from both data and expert opinion. The principal advantages over statistical methods is the ability to examine the causes driving certain outcomes (e.g., ruin scenarios), and the ability to directly model the effect of different decisions on the outcome.

- Dynamic Financial Analysis (DFA) — the name for a class of structural simulation models of insurance company operations, focusing on certain hazard and financial risks and designed to generate financial pro forma projections.

*Note: As a practical matter and as noted above, the choice of modeling approach is typically between **statistical analytic models** and **structural simulation models**. The contrast between these modeling approaches is summarized in the table below:*

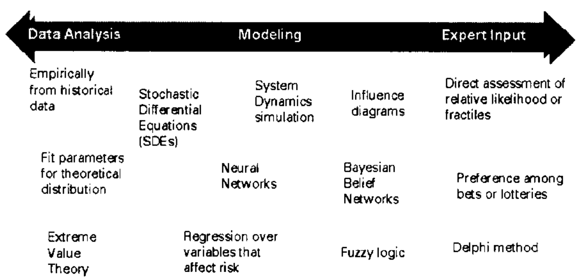
Representation of Relationships	Calculation Technique	Examples	Relative Advantages
Statistical (based on observed statistical qualities without regard to cause/effect)	Analytic (closed-form formula solutions)	▪ RBC ▪ Rating agency models	Speed; ease of replication; use of publicly available data (well suited for industry oversight bodies)
Structural (based on specified cause/effect linkages; statistical qualities are outputs, not inputs)	Simulation (solutions derived from repeated "draws" from the distribution)	▪ DFA ▪ Many options pricing models	Flexibility; treatment of complex relationships; incorporation of decision processes; ability to examine scenario drivers (well suited for individual companies)

The models described above generally presuppose the existence of sufficient data with which to fully parameterize the models. This is often not the case in practice, particularly as respects operational and strategic risks.

There is a wide variety of risk modeling methods that can be applied to a specific risk. They can be thought of as lying on a continuum that is based on the extent to which they rely on historical data vs. expert input (see Figure A below). Along the continuum of sources of information, the methods listed on the left are ones that rely primarily on the availability of historical data. They include, for example, empirical distributions, parametric methods to fit theoretical probability density functions, regression, stochastic differential equations and extreme value theory. These methods have been used extensively by financial institutions to model financial risks.

The methods listed on the right in Figure A rely primarily on expert input, including for example, Delphi method, preference among bets or lotteries, and influence diagrams. These have been used successfully for several decades by decision and risk analysts to model operational risks in support of management decision-making in manufacturing, particularly in the oil and gas industry, and in the medical sector. The methods listed in the middle of the continuum rely on data, to the extent that it is available, and expert judgment to supplement the missing data. In these methods, expert judgment is used to develop the model logic indicating the interactions among key variables and to quantify cause/effect relationships based on experience and ancillary or sparse data. Methods such as system dynamics simulation, Bayesian belief networks and fuzzy logic in particular are ideally suited for quantifying operational and strategic risks.

Figure A – There is a continuum of methods for modeling risks. Each method has advantages/disadvantages over others, so it's important to select the best methods based on facts and circumstances



Definitions and descriptions of the risk modeling methods that lie along this continuum are in Appendix B.

Risk Integration

Several of the risks of interest to the organization may be correlated with one another. For example, economic inflation (a driver of cost trends across multiple business segments) is highly correlated with interest rates (a driver of asset values and investment returns). It is important to capture these correlations – indeed, this is the essence of ERM. There are several ways to do this.

A direct way to express dependencies among risks is to estimate the statistical correlations between each of the individual risks. These estimates are often arrayed in a "covariance matrix".

- Covariance — a statistical measure of the degree to which two random variables are correlated. Related to correlation coefficient (correlation coefficient = covariance divided by the product of the standard deviations of the two random variables). A correlation coefficient of +1.0 indicates perfect positive correlation; -1.0 indicates perfect negative correlation (i.e., a "natural hedge"); zero indicates no correlation.

- Covariance matrix — a two-dimensional display of the covariances (or correlation coefficients) among several random variables; the covariance between any two variables is shown at their cross-section in the matrix.

The estimation of these covariances can be a practical difficulty, as the number of estimates required rises as the square of the number of risks.

An alternative way to capture risk interrelationships is through a *structural simulation model* of the enterprise, described above. In essence, a structural simulation model allows one to capture the dependencies among variable inputs in a simple, accurate and logically consistent way by virtue of the model's cause/effect linkages of these inputs to common higher-level inputs.

For example, interest rates and inflation rates are often generated stochastically by means of an economic scenario generation model, wherein these two random variables are linked to higher-level economic forces. In turn, other lower-level random variables, such as product costs, prices, asset values and investment income, are linked causally to interest rates and inflation rates within the model. Without such structural linkages, other models (such as MVC models, described above) can generate sets of random variables that are unrealistic relative to each other, regardless of how accurate the correlation estimates among them may be.

The statistical correlations among risks that are related through a structural simulation model are an emergent property (i.e., an output) of the model, not values to be separately estimated. To the extent that certain inputs are not related to a common higher-level input, yet one believes that a relationship exists between them, these correlations can be stated explicitly in terms of a covariance matrix, whose values can be determined through data analysis, expert opinion or both.

Risk Prioritization

Risk prioritization is ranking material risks on an appropriate scale, such as frequency, severity or both.

- Risk mapping — the visual representation of identified risks in a way that easily allows ranking them. This representation often takes the form of a two-dimensional grid with frequency (or likelihood of occurrence) on one axis, and severity (or degree of financial impact) on the other axis; the risks that fall in the high-frequency/high-severity quadrant are typically given highest priority risk management attention.

A more useful ranking of risks is in terms of each risk's impact on the organization's overall key performance indicators (KPIs). The marginal contribution of each individual risk factor to the overall risk profile of the organization can be determined by "turning off" that risk factor (changing that particular input from stochastic to deterministic) and examining the impact on the KPI probability distribution. This technique provides a straightforward way of isolating the impact of a particular risk factor (such as natural catastrophes) on overall capital adequacy, for example. In this way, the prioritization of risk factors, which is often done qualitatively, can be more rigorously validated.

Tool Applications for Treating/Exploiting Risks

The techniques, models and measures above are used in various combinations to assist management decision-making in a number of areas. Several of these specific applications are discussed below, following the definitions of two generic applications ("optimization" and "candidate analysis") that are employed within some of these specific applications. Note that the following list of specific applications is not exhaustive, and is expected to grow as ERM matures as a discipline. Virtually any decision that requires evaluating risk/return trade-offs is a candidate for ERM treatment.

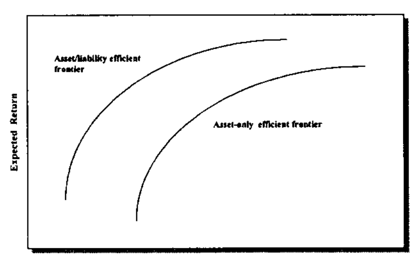
- Generic applications:
  - Optimization — the formal process by which decisions are made under conditions of uncertainty. Components of an optimization exercise include a statement of the range of decision options, a representation of the uncertain conditions (usually in the form of probability distributions), a statement of constraints (usually in the form of limitations on the range of decision options), and a statement of the objective to be maximized (or minimized). An example of an optimization exercise is an asset allocation study (see below under risk management applications). [See also "candidate analysis, below.]
  - Candidate analysis — a restricted form of optimization analysis in which only a finite number of prespecified decision options are considered, and the best set among those options is determined through the analysis. Optimization and candidate analyses can be contrasted as follows. An optimization analysis would typically result in the derivation of an "efficient frontier" curve in risk/return space, which contains the decision options that result in maximum return for each level of risk (i.e., the optimal decision option for each level of risk). A candidate analysis would not derive the efficient frontier curve, but would simply show the finite number of decision options in comparison with each other in risk/return space (i.e., a "scatter plot"). It would not be known how close each option is to the efficient frontier of options. Conceptually, if a candidate analysis were performed on an infinite number of candidate decision options, then the "envelope" or boundary of those options would form the efficient frontier.

- Capital management:
  - Capital adequacy — the determination of the minimum amount of capital needed to satisfy a specified economic capital constraint (e.g., a certain probability of ruin), usually calculated at the enterprise level.

- Capital structure — the determination of the optimal mix of capital by type (debt, common equity, preferred equity), given the risk profile and performance objectives of the enterprise.
- Capital attribution — the determination of the assignment of enterprise level capital to the various business segments (e.g., lines of business, regions, projects) that make up the enterprise, in recognition of the relative risk of each segment, for purposes of measuring segment performance on a risk-adjusted basis (e.g., to provide the denominator for a RORAC or RARORAC analysis by segment).
  - Diversification credit — the recognition of the "portfolio effect", which is the fact that the economic capital required at the enterprise level will be less than the sum of the capital requirements of the business segments calculated on a stand-alone basis. The diversification credit is typically apportioned to the business segments in a manner that attempts to preserve the relative equity of the capital attribution process.
- Capital allocation — the actual deployment of capital to different business segments.

- Performance measurement — the development and implementation of appropriate risk-based metrics for evaluation of business segment performance, reflecting capital consumption, return and volatility.

- Investment strategy/asset allocation — the determination of the optimal mix of assets by asset class (usually to maximize expected return at each level of risk, i.e., according to Modern Portfolio Theory). In advanced applications, the analysis reflects the nature and structure of both assets and liabilities and is called asset/liability management (ALM).



- Insurance/reinsurance/hedging strategy optimization — the determination of the optimal insurance/reinsurance/hedging program, reflecting program costs and risk reduction capability; usually conducted through candidate analysis. The risk reduction capability manifests itself in terms of both reduction in required economic capital and reduction in the cost of capital or required risk-adjusted rate of return.

- Crisis management — the proactive response of an organization to a severe event that could potentially impair its ability to meet its performance objectives.

- Contingency planning — the process of developing, and embedding in the organization, crisis management protocols in advance of crisis conditions.

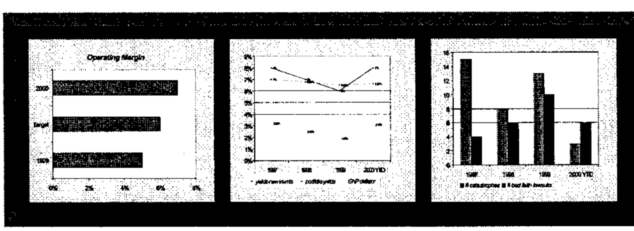
- Business expansion/contraction strategy — the evaluation of merger, acquisition and divestiture options in terms of their incremental impact on the risk profile of the enterprise.

- Distribution channel strategy — the systematic evaluation of alternative channels (e.g., direct, agency, Internet), by means of simulation analysis to test impacts on growth, market share, profitability, etc. on a risk/return basis.

- Strategic planning — the use of structural simulation modeling, such as "real options" modeling, as a decision tool to assist management in selecting among alternative strategies, such as long-term research projects (see "Scientific Management at Merck", *Harvard Business Review*, 1994).

Risk Monitoring

Continual monitoring of the risk environment, and of the performance of the risk management processes, is often done by means of a senior management *risk dashboard* — the graphical presentation of the organization's key risk measures (often against their respective tolerance levels), as in the chart below.



Typical measures included in the dashboard are shown in the following tables.

Marketing

- New business sold
- Retention of old business
- Mix of business: new and renewal
- Market share by customer type
- Average premium or assets by per customer
- % high-yield customers
- Customer satisfaction
- Average # of products per customer

Underwriting

- Price achieved vs. target price
- Exposure data (number of cars, payroll, etc.)
- Exposure mix
- Quotes accepted/declined
- Variance analysis
- Premium persistency
- Loss ratio
- Loss adjustment expense

Financial

- Revenue
- Underwriting profit
- Investment profit
- Pre-tax operating income
- Net income
- Return on equity and total capital
- Economic value added

Sales/Distribution

- Acquisition costs per sale
- Sales by distribution channel
- Growth/retention of agents

Investments

- Cash flow
- Yield on new investments
- Yield on portfolio by class and duration
- Convexity of assets
- Duration of assets
- Investment mix: new and portfolio
- Credit default
- Total return

Human Resources

- Agency composition (number, age, service)
- Total employment by department
  - Number and percentage leaving the company
  - Vacancy rates
  - Average salary increase vs. plan
- Employee commitment and engagement

Claims

- Frequency and severity of claims
- Claims department productivity

External Data

- Audit compliance
- Inflation rates
- Interest rates
- GNP
- Competitor pricing

* * * * *

Note: Certain material in this chapter was drawn from the article "The Language of Enterprise Risk Management: A Practical Glossary and Discussion of Relevant Terms, Concepts, Models and Measures", by Jerry Miccolis, in the Enterprise Risk Management Expert Commentary section of the Web site of the International Risk Management Institute, http://www.irmi.com/expert/risk.asp. As noted therein, certain of these definitions were adapted from *The Dictionary of Financial Risk Management*, by Gastineau and Kritzman, 1996, Frank J. Fabozzi Associates. Certain other material was drawn from the Tillinghast – Towers Perrin monograph *RiskValueInsights™: Creating Value Through Enterprise Risk Management*, (http://www.tillinghast.com).

V. ERM Case Studies

This chapter recounts a number of success stories in which organizations made the commitment to and then benefited from ERM. Some of these benefits are explicit and measurable (e.g., increased investment returns, decreased capital requirements), others are more intangible but no less real (e.g., more enlightened strategic planning, more rigorous performance measurement/management). There should be elements from this collection of cases that will resonate with any given organization.

It also should be clear from these cases that, in terms of objectives, scope (of risks and of processes), organization, tools and techniques, there are a number of legitimate approaches to ERM and no single "correct way" that is appropriate for all entities. The proper approach to ERM for any enterprise is one that fits within the culture of that enterprise.

Risk Assessment

A large, market-leading manufacturer and distributor of consumer products with an uninterrupted 40-year history of earnings growth, embarked on ERM well before its competitors. This step followed their philosophy of "identifying and fixing things before they become problems". They were spurred by their rapid growth, increasing complexity, expansion into new areas, and the heightened scrutiny that accompanied their recent initial public offering. They conducted a comprehensive assessment of all risks that could potentially prevent the company from achieving its promised results. Views of company executives on key performance measures and risk thresholds were validated against financial models of stock analyst expectations. Multiple methodologies were used to rank order risks from all sources (hazard, financial, operational and strategic) on the basis of expected impact, and the results cross-validated. High-priority risk factors were interpreted and classified (as "strategic", "adaptation", "manageable", "business as usual") for appropriate response, and strategies for mitigation and exploitation were developed. In addition, a "Business Risk Self-assessment Toolkit" was created for ongoing use. Senior management attributes the ERM effort, and their communication of that effort to the investment community, as one of the drivers of the company's superior market valuation.

A large health plan had traditionally conducted separate and uncoordinated risk assessments through its risk management, legal and internal audit functions. It undertook an enterprise-wide risk assessment covering all functional and operational divisions. The objective was to prioritize all sources of risk against a common set of financial and customer metrics to enable senior management to focus the organization's limited resources on the proper short list of critical concerns. In addition to providing a meaningful and useful calibration of risks of varied types, this exercise surfaced critical business risks that had not been identified through any previous audit or strategic planning exercise. Senior management uses the results of this assessment to set its strategic agenda.

Distribution Strategy

A medium-sized life insurance company wanted to reconsider their distribution strategy in light of plans to demutualize the following year. The bulk of their production came from a network of career agencies, and the company wanted to investigate not only other distribution channels but also the possibility of becoming a wholesaler to other financial institutions. They decided to analyze the risk/value economics of alternative operational strategies by developing a financial model of the underlying business dynamics. The process of model development and assumption setting forced the management team to articulate the alternative strategies more clearly and with greater specificity than they had thus far. The model was used iteratively to evaluate further variations in strategy suggested by a review of the projected financials at each prior iteration. Modeling the economics provided the management team with valuable information on the risks and opportunities underlying alternative strategies. As a result, the team was able to reach consensus on a distribution strategy that was better understood and provided the best prospects of success.

Performance Measurement

A large multinational financial services group undertook an assessment of the relative levels of economic capital required by each of its life and nonlife insurance subsidiaries. This involved identifying the major sources of risk in each line of business and modeling the impact of these risk areas on the projected cash flows. The results were used to determine an appropriate level of capital at individual product level, subsidiary level, product group level (across subsidiaries) and finally at group level. An economic scenario generation model was used to allow cross-currency aggregation. The resulting attribution of capital is used as the foundation for a performance measurement system relating shareholder risk to return on capital and total shareholder return. Actual return on capital is compared to the hurdle rate implied by the shareholder risk and differences are analyzed into above- and below-the-line effects.

Asset Allocation

A property/casualty insurance company's conservative asset mix resulted in performance returns that were not competitive. They evaluated alternative asset allocation strategies, along with an integrated reinsurance program, to enhance the returns from investments and manage the risk of their business. However, the company did not want its rating from A.M. Best to be affected as a result of implementing a more aggressive investment strategy. They developed a comprehensive model of the company and evaluated multiple scenarios of economic value in relation to risk. The model allowed them to develop a strategy to alter their asset allocation. A financial integrated stop-loss reinsurance program was designed with an investment hedge to mitigate the possibility that the investment portfolio may underperform a target return. The result: enhanced expected returns of the investment portfolio and lowered downside risk on operating income. The executive team's understanding of their return opportunities in relation to the risks of the business was deepened. This insight was used to focus the work of line managers, and also used in discussions with outside parties regarding overall risk management.

Strategic Planning

A leading global manufacturer and distributor of patented pharmaceuticals has developed its ERM approach around a "real options" model. In an industry noted for very expensive, very long-term research projects, success is dependent on making the right "bets" on those research projects, both at their outset and at critical decision junctures throughout the projects' life span. The company credits its pioneering work on its Research Planning Model as a key contributor to its competitive advantage. This model captures the important medical, operational and financial risks of each project, and applies sophisticated options pricing theory to discern among alternative projects and to manage the continuing investments in projects that pass the initial screening process. This approach, by recognizing the dynamics of the *staged* research decision process, has allowed the company to pursue ultimately successful projects that would have failed a more traditional net present value screening process. (*Note*: This case study is documented in "Scientific Management at Merck: An Interview with CFO Judy Lewent", *Harvard Business Review*, January-February 1994.) Certain tools developed for this approach – most notably "decision trees" – have become routinely used in management discussions of unrelated issues throughout various organizational levels, thus contributing to the company's "common language of risk".

The board of directors at a large electric utility, motivated both by local corporate governance guidelines and the opening of their industry to competition, mandated an integrated approach to risk management throughout the organization. They piloted the process in a business unit that was manageable in size, represented a microcosm of the risks faced by the parent, and did not have entrenched risk management systems. This same unit was the focus of the parent's strategy for seeking international growth – a strategy that would take the organization into unfamiliar territory – and had no established process for managing the attendant risks in a comprehensive way. The pilot project was deemed a success and, among other things, the ERM unit is now a key participant in the organization's strategic planning process. This participation takes the form of building stochastic models around the key drivers of the strategic plan (weather conditions, customer demands, economic conditions, etc.) to assess the robustness of the plan. The board will not approve the strategic plan without such an ERM evaluation.

Product Design

A life insurer was looking to improve the product design features of its flagship universal life product; specifically, incorporating a market value adjustment to protect against having to credit high interest in times of falling asset market values. The market value adjustment could have been a serious detriment to potential policyholders and might not have received regulatory approval. Working together, senior management, an actuarial team and the investment fund manager determined that an ALM model be developed using a set of stochastically generated interest rate scenarios. Various investment strategies were considered, covering a varying mix of mortgages, high-quality corporate bonds and CMO's. The ALM model then made projections based on the modeled relationship between the yield on these asset classes and the yield curve for treasuries as produced by the stochastic interest rate generator. Appropriate assumptions were made for defaults and prepayment risk. The yield relationships and other asset assumptions were reviewed by the fund management team, which also appraised the actuaries' assumptions underlying the model that was used to create the stochastically generated interest rate scenarios. Duration and convexity of both assets and liabilities were then analyzed, and the product design and the planned investment strategy fine-tuned to bring the assets and liabilities into balance. At this point, senior management analyzed various profit metrics for different investment strategies, looking at extreme scenarios for special review. Based on this analysis, the product appeared to hold up well even under the most extreme interest rate scenarios without any market value adjustment. The ALM analysis was effectively used to establish the product design and set the investment policy, and the product was filed without any market value adjustment.

Dividend Strategy

A medium-sized foreign life insurance company wanted to analyze the viability of their current dividend strategy for traditional business. Its market provided stable long-term dividend rates at a high level, even while market interest rates have declined, by smoothing book yields via accrual and realization of "hidden" reserves (unrealized capital gains on assets) and unallocated bonus reserves. In the prevailing low interest rate environment, the key competitive issue had become how long companies could finance their current dividend rates from existing buffers as compared to the market. In order to analyze the company's competitive position, ALM models were built for the company and a representative market company, reflecting the company's specific portfolio structure and strategies. On the basis of stochastic scenarios generation, the estimated time until ruin (until buffers had been exhausted) was determined for a range of potential ALM strategies for the company and compared to the results for the market. By varying the investment strategy, the company improved its risk/return positioning. As a result of the benchmark study, the life insurer received an indication of its current competitive position and a quantification of alternative ALM strategies, which led the company to reassess its dividend setting strategy for the entire traditional life portfolio.

Risk Financing

A very large retail company's CFO wanted to "assess the feasibility of taking a broader approach to risk management in developing the organization's future strategy". As part of this effort, they hoped to "evaluate our hazard risk and financial risk programs and strategies, to identify alternative methods of organizing and managing these exposures on a collective basis". As a first step, the company designed and built a model to provide an improved capability to evaluate its hazard and financial risks, both individually and on an aggregate portfolio basis. Criteria were developed to evaluate alternative risk financing programs based on appropriate measures of performance for risk and return. These evaluation criteria allowed the company to develop risk/return "efficient frontiers", representing a range of possible changes from their current program, on which to make informed management decisions. These decisions included:

- Choosing among competing insurance program submissions
- Determining retention levels
- Developing negotiating strategies
- Designing an overall risk financing strategy
- Prioritizing risk management activities (e.g., risk control).

The process for developing this capability included the determination of both appropriate return measures (e.g., net income, net cash flow) and appropriate risk measures (e.g., magnitude of potential loss, variance in financial measures, liquidity, compliance with bond covenants). These measures recognized and were developed from the variety of needs of key decision-makers, identified via structured interviews. Additionally, the process provided an understanding of those factors that have the greatest impact (in risk and return terms) on the performance of individual risks as well as the portfolio of all risks. To codify this process, the company developed a computer-based decision-support tool (with "senior management-friendly" graphics) that facilitated the evaluation of hazard and financial risks and allowed the decisions to be fact-based and consistent.

* * * * *

In addition to these examples, there are numerous others that demonstrate additional collateral benefits to undertaking an ERM process. These include:

- Improved communication and collaboration within the organization;

- Better-informed decisions at all levels in the organization by having gone through a rigorous and systematic risk identification/prioritization process; and

- Valuable change in mindset wherein risk can be a source of opportunity and not merely a threat to be avoided.

VI. Practical Considerations in Implementing ERM

Once an enterprise decides to adopt ERM, it has to deal with a number of practical considerations in its successful implementation. These include, but are not limited to, the following:

Designating an ERM "Champion"

Given the implementation challenges, a unique individual is needed to spearhead the effort, becoming, in effect, the "champion" of the initiative. This role is often fulfilled by naming a Chief Risk Officer (CRO), who typically reports to the Chief Executive Officer or Chief Financial Officer. It is important that the organizational structure created for ERM (e.g., the CRO, the CRO's staff, the Risk Management Committee) is accountable and has the authority to be a change agent. Senior sponsorship needs to be high enough in the organization to have a top-level view of all the risks facing the enterprise, see across all organizational "silos", and have sufficient authority to effect changes in business practice.

Making ERM part of the enterprise culture ("tearing down the silos")

Under the historical, fragmented approach to risk management, numerous personnel are involved in various aspects of risk management. Typical of such approaches, the risk management department is responsible for hazard risks; the treasury department is responsible for financial risks; the human resources department is responsible for workers compensation, health, and employee risks; information technology is responsible for many operational risks; and the marketing department is responsible for many strategic risks. More than likely, these departments report to different managers within the organization, use different risk assessment procedures and terminology, calibrate risk on different scales, and have different timeframes in mind. Instituting such a sweeping change as implementing ERM may invoke defensive postures as these departments try to protect "their turf". The successful ERM approach would be one that coordinates all these different departments, recognizes the need for education, but allows for individual department initiative, flexibility, and autonomy.

Determining all possible risks of the organization

As the list of risks included in the ERM Framework demonstrates, there is a multitude of risks facing every enterprise. Often the greatest risks are those not contemplated. Who in the property and casualty insurance industry could have conceived the magnitude of environmental risks assumed in insurance policies prior to the mid-1980's, or the terrorism exposure in the early 2000's? Who in the pharmaceutical industry could have conceived of effect of criminal tampering with products on store shelves? How can these risks be quantified, integrated or treated, if they cannot be identified? Some organizations have used their risk management committees to conduct and participate in periodic, structured "disaster scenario" brainstorming exercises specifically to contemplate and, as appropriate, plan for such "unthinkable" events.

Quantifying operational and strategic risks

Although a great body of literature exists in the quantification of hazard and financial risks, not all enterprises are able to quantify intangible risks such as operational and strategic risk. It is difficult to determine point estimates of likelihoods (i.e., frequency) and consequences (i.e., severity) of these risks, let alone determine probability distributions around these estimates. Not only do models generally not exist, but historical data that are the input to these models often do not exist either. Even if attempted, the cost of quantifying these risks needs to be considered in relationship to its benefit.

Enterprises can overcome these difficulties by starting with qualitative analysis of operational and strategic risk to determine those that are material and to prioritize them. In addition, some have advocated the use of causal models, as opposed to parametric models, to quantify these risks. These causal models often already exist (e.g., in strategic planning, in logistics) in some form within the organization and may simply need to be "stochasticized".

Integrating risks (determining dependencies, etc.)

Actuaries and financial analysts know of the difficulty in determining appropriate relationships or correlations for risks just within their respective areas of expertise, hazard and financial risks. These difficulties include:

- Past causal relationships are often not indicative of future relationships.
- There are differences in time frames (short-term, medium-term, long-term) to consider.
- Selecting correlation factors becomes cumbersome as the number of risks to review increases.

These difficulties are compounded when considering operational and strategic risks, both within these risk categories and among other risk categories.

Building structural models in modular form, which allows enhancement in manageable successive stages over time, is one practical approach some companies have employed.

Lack of appropriate risk transfer mechanisms

Although risk transfer mechanisms for hazard and financial risks exist via the insurance, reinsurance and capital markets, these markets are not complete in the sense of being able to provide all products and services that enterprises may need. These markets need to continue to evolve over time (such as the development of the alternative risk market for hazard risks) in order to provide products that will meet the risk transfer needs of enterprises. Risk transfer mechanisms for operational and strategic risks are even less mature.

Monitoring the Process

Ideally, ERM is not a one-time "project", but a discipline that evolves over time as risks and opportunities within an enterprise change. The successful ERM process will include regular progress reports and comparisons to previous risk assessments so changes and refinements can be made as appropriate. Changes in the risk environment, based on new information, may result in changing strategies employed to treat and exploit risk. Regularly monitoring results can, and should, be tied to the time scales identified for the risks actively managed.

Start Slowly – Build Upon Successes

Because of the traditional, fragmented approach to risk management described earlier and the complexity of many businesses, enterprises often find it useful to start their ERM initiative slowly, tackling smaller projects first, so tangible results can be achieved early. The CRO or Risk Management Committee or both also may have limited resources initially, so they have to think on a smaller scale until successful projects are completed. However, the early successes can help to generate momentum and enthusiasm (and perhaps funding) for future ERM initiatives.

The case studies in the preceding chapter include examples of how different companies in various industries started small in terms of any or all of the following:

- Risk type (e.g., combining hazard and financial risks first, then planning to layer in strategic and operational risks);
- Process step (e.g., starting with a qualitative enterprise-wide risk assessment, then proceeding to risk quantification);
- Organizational component (e.g., piloting ERM within a single corporate division).

Just as there is no one correct approach to overall ERM design, there is no one correct path to incrementally building toward ERM. Both are dependent on the unique business imperatives and culture of each organization.

Appendix A — Risk-Related Regulatory, Rating Agency and Corporate Governance Guidelines and Requirements

Those developing ERM programs and policies need to consider a number of corporate governance guidelines and regulatory and rating agency requirements. The more prominent of these are described below.

- General Industry
  - Cadbury Report, et al (U.K.) — the London Stock Exchange has adopted a set of principles, the Combined Code, that consolidates previous reports on corporate governance by the Cadbury, Greenbury and Hampel Committees. This code, effective for all accounting periods ending on or after December 23, 2000 (and with a lesser requirement for accounting periods ending on or after December 23, 1999), makes directors responsible for establishing a sound system of internal control and reviewing its effectiveness, and reporting their findings to shareholders. This review should cover all controls, including operational and compliance controls and risk management. The Turnbull Committee issued guidelines in September 1999 regarding the reporting requirement for non-financial controls.
  - Dey Report (Canada) — commissioned by the Toronto Stock Exchange and released in December 1994, it requires companies to report on the adequacy of internal control. Following that, the clarifying report produced by the Canadian Institute of Chartered Accountants, "Guidance on Control" (CoCo report, November 1995), specifies that internal control should include the process of risk assessment and risk management. While these reports have not forced Canadian listed companies to initiate an ERM process, they do create public pressure and a strong imperative to do so. In actuality, many companies have responded by initiating ERM processes.
  - Australia/New Zealand Risk Management Standard — a common set of risk management standards issued in 1995 that call for a formalized system of risk management and for reporting to the organization's management on the performance of the risk management system. While not binding, these standards create a benchmark for sound management practices that includes an ERM system.
  - KonTraG (Germany) — a "mandatory bill" that became law in 1998. Aimed at giving shareholders more information and control and increasing the duty of care of the directors, it includes a requirement that the management board establish supervisory systems for risk management and internal revision. In addition, it calls for reporting on these systems to the supervisory board. Further, auditors appointed by the supervisory board must examine implementation of risk management and internal revision.

- Financial Services Industry
  - Basel Committee:
    - The Basel Committee on Banking Regulation and Supervisory Practices was established in 1974 (originally called the Cooke Committee) in response to the

erosion of capital in leading global banks. The committee meets under the auspices of the Bank for International Settlements (BIS) but is not part of the BIS. The committee consists of representatives from the central banks/supervisory authorities of the G10 countries and Luxembourg. The committee has no legal authority, but the governments of the representatives on the committee have always legislated to make the recommendations part of their own national law. The standards set by the committee are widely regarded to be best practice and a large number of other countries that are not formally represented on the committee have implemented the proposals. In the U.S., the Federal Reserve has adopted the Basel Capital Accord ("Basel I" – see below).

— "Basel I" — the 1988 Basel Capital Accord established a framework to calculate a minimum capital requirement for banks. The Accord focused on credit risk and was crude in its recognition of the relative risk of different loans. A number of amendments were made to the Accord (prior to "Basel II" – see below), the most significant of which is the market risk amendment in 1996; this extended the 1988 Accord to cover market risk and allowed for the use of internal models to quantify regulatory capital.

— "Basel II" — in 1999 the Basel Committee issued a draft proposal for a new accord and accepted comment. Based on feedback, the Committee issued a revised proposal in 2001 for review and comment. In this New Basel Capital Accord, proposed for implementation in 2004, among other changes a capital charge for operational risk is included as part of the capital framework. The charge reflects the Committee's "realization that risks other than market and credit" can be substantial. Operational risk is defined as "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events". The new capital adequacy framework is proposed to apply to insurance subsidiaries of banks and may apply to insurance companies as insurance and banking activities converge.

- OSFI (Canada) — the Office of the Supervisor of Financial Institutions supervisory framework defines "inherent risk" to include credit risk, market risk, insurance risk, operational risk, liquidity risk, legal and regulatory risk and strategic risk. It states that: "Where independent reviews of operational management and controls have not been carried out or where independent risk management control functions are lacking, OSFI will, under normal circumstances, make appropriate recommendations or direct that appropriate work be done."
- FSA (U.K.) — the Financial Services Authority (FSA – the recently created regulator of all U.K. financial services businesses) is introducing a system of risk-based supervision that will create a single set of prudential requirements organized by risk rather than by type of business. Regulated businesses will have to demonstrate that they have identified all material risks and have adequate systems and financial resources to manage and finance such risks, including market risk, credit risk, operational risk and insurance risk. There is also likely to be a requirement for formal documentation of the whole process in a format that is readily accessible to the FSA.

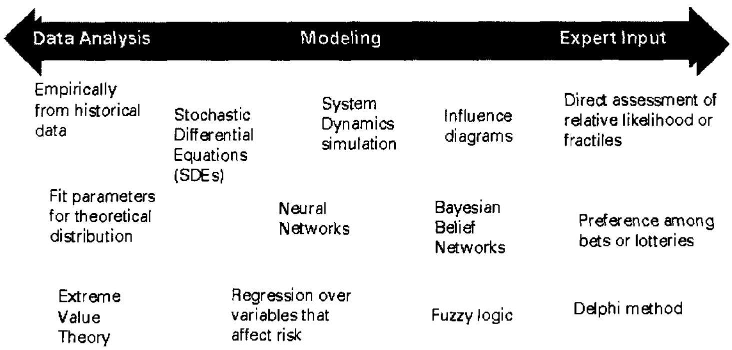
- Insurance Industry
  - A.M. Best — in its publication *Enterprise Risk Model: A Holistic Approach to Measuring Capital Adequacy*, A.M. Best describes its VaR-based method for determining the adequacy of capital for rating purposes. The report states: "The Enterprise Risk Model is a modular system designed to capture all risks, including noninsurance and non-U.S. related risks. VaR methodologies are somewhat controversial in insurance circles, but they are the standard for other financial-services organizations. More importantly, A.M. Best believes that VaR-based methodologies provide a more accurate assessment of risk and required capital, since they use observable market metrics. Beyond its application in the rating process, the model can also be a useful tool for financial managers, since the VaR framework provides a natural springboard to other applications, including risk-adjusted return on capital (RAROC) and dynamic financial analysis (DFA). The Enterprise Risk Model quantifies the risk to the future surplus – net worth – of an organization arising from a change in underlying risk variables, such as credit risk, insurance risk, interest rate risk, market risk and foreign exchange risk. The model also quantifies the benefits of diversification as it takes a macro view of the correlations among risks within an organization...Like other VaR-based models, it is calibrated to measure the risks over a defined holding period – one year – for a given level of statistical confidence – 99%."
  - Moody's — in its publication *One Step in the Right Direction: The New C-3a Risk-Based Capital Component* (June 2000), Moody's Investors Service states that it will use the new method devised by the NAIC and the American Academy of Actuaries for measuring a life insurance company's C-3a (interest rate) risk, as it incorporates a cash-flow testing requirement for annuity and single premium life products and is more consistent with industry advances in dynamic cash-flow testing. *One Step* states: "...the revised calculation is a more accurate barometer of the amount of capital required to support an insurer's interest-sensitive business, as it explicitly incorporates asset-liability mismatches in determining the appropriate amount of required regulatory capital for a company. Consequently, the new calculation should help discourage companies from taking unwarranted asset-liability risk."
  - Standard &Poor's — in its *Revised Risk-Based Capital Adequacy Model for Financial Products Companies* Standard & Poor's states: "Standard & Poor's Insurance Capital Markets Group has developed a new, risk-based capital adequacy model to analyze the credit, financial market, and operational risks of companies that are offering products or are using sophisticated risk management techniques that are not considered under the existing Rating Group's capital models. The model will also determine these companies' capital adequacy. The primary application of the model will be to analyze specialized financial product companies (FPCs) that are subsidiaries of insurance companies or that are credit enhanced by insurance companies...The model may also be applied to portions of insurance companies that control or mitigate their risks to a greater extent than is implied by the capital charges applied in the standard life/health capital adequacy

model, which bases charges for interest-rate risk and credit risk on industry averages and liability types rather than company-specific exposure."

- NAIC — The National Association of Insurance Commissioners:
  - Risk-Based Capital (RBC) — Following a detailed examination of the growing diversity of business practices of insurance companies conducted in 1990, the NAIC concluded that minimum capital requirements placed on companies needed to be increased to protect consumers. The NAIC adopted life/health risk-based capital requirements in December 1992 and adopted property/casualty risk-based capital requirements in December 1993. Although risks involved in these two segments of the industry are very different, the NAIC was able to develop a consistent two-step approach to setting risk-based capital requirements for individual companies:
    - Step 1 involves the calculation of a company's capital requirement and total adjusted capital, based on formulas developed by NAIC for each industry.
    - Step 2 calls for comparison of a company's total adjusted capital against the risk-based capital requirement to determine if regulatory action is called for, under provisions of the Risk-Based Capital for Insurers Model Act. The model law sets the points at which a commissioner is authorized and expected to take regulatory action.
  - Interest rate risk — the NAIC's Life Risk-Based Capital Working Group, in conjunction with the American Academy of Actuaries Life Risk-Based Capital Task Force, has finalized the development of an improved method for measuring a company's interest-rate risk. The method, which is effective for the year-end 2000 statements, "incorporates a cash-flow testing requirement for annuity and single premium life products and makes the RBC C-3a calculation more consistent with recent industry advances in dynamic cash-flow testing...The task force has recognized the need to accurately incorporate these additional risks into the RBC formula. They have stated that equity indexed annuities (EIAs) and variable products with secondary guarantees will be incorporated in a future C-3a update. This would be consistent with the task force's goal of upgrading C-3a from a measure of interest-rate risk to a more complete measure of asset/liability risk."
- Australian Prudential Regulation Authority (APRA) — a feature of ongoing reforms to the regulation of general insurers is a layer of four standards covering the subjects of capital adequacy, liability valuation, reinsurance arrangements and operational risk. APRA is implementing an approach based on development of, and compliance with, a range of risk management strategies. These strategies will need to deal with the myriad interlocking risks involved in managing a general insurance company. Each company will need to have its strategy agreed upon by APRA and will then be responsible for managing compliance. APRA has made it clear that an internal enterprise risk model with appropriate specifications will go a long way toward meeting compliance objectives.

Appendix B — A Continuum of Risk Modeling Methods

Figure A – There is a continuum of methods for modeling risks. Each method has advantages/disadvantages over others, so it's important to select the best methods based on facts and circumstances



There is a continuum of methods for developing probability distributions. The choice of method depends significantly on the amount and type of historical data that is available. The methods also require varying analytical skills and experience. Each method has advantages and disadvantages over the other methods, so it is important to match the method to the facts and circumstances of the particular risk type.

We have loosely organized the modeling methods into three categories:

- Methods based primarily on analysis of historical data
- Methods based on a combination of historical data and expert input
- Methods based primarily on expert input

Methods Based Primarily on Analysis of Historical Data

These methods are the most appropriate when there is enough historical data to apply standard statistical approaches to develop probability distributions. Typically several years of high-frequency data are necessary. These methods are most often used to model risks that are traded in the financial markets such as interest rate, foreign exchange, asset risks, claims and the like.

Empirical Distributions

The simplest and the most direct approach is to assume that the historical data fully defines the probability distribution. Then the data can be used directly to develop a discrete probability distribution. Of course the danger is in assuming that the data is complete and the time period over which the data is gathered is long enough to have "seen" or experienced the full range of outcomes.

Fit Parameters of Theoretical Probability Density Functions

An alternative to empirical distributions is to assume that the risk can be described by a theoretical probability density function. Then the data is used to estimate the parameters of the theoretical distribution. For example, for property/casualty claims, the frequency of claims is often assumed to follow either a Poisson or negative binomial distribution whereas the severity of claims is often assumed to follow a lognormal or a Pareto (for conditional claim or tail distribution).

Stochastic Differential Equations (SDE)

A Stochastic Differential Equation (SDE) expresses the difference (or change) in the value of a variable (e.g., interest rate) at time t and the value one time period later, t +1. It's a **stochastic** differential equation because the difference is expressed as a combination of a predictable change and an uncertain or random change during the time period. The random change is represented as a random variable with a specified probability distribution (typically normal distribution). Starting with an initial value, the SDE is used to iteratively determine a scenario of how the value changes over a forecast period (e.g., 10 years). Hundreds or possibly thousands of scenarios are generated in this way. The scenarios can then be summarized as probability distributions for each point in time over the forecast period. See the ERM bibliography for helpful publications that provide more detail on use of SDEs to model risk.

Extreme Value Theory

In risk management, often the most important part of a probability distribution is the tail representing the downside risk. The tail distribution is used to determine capital and shortfall risk constraints for optimizing strategies. However, most risk modeling methods focus on accurately representing the main body of the distribution. Extreme Value Theory (EVT) is a technique for increasing the accuracy with which to model the probability of large values in the tail distribution. EVT is devoted to the modeling and estimating the behavior of rare events. Different EVT models and techniques have been developed and applied to deal with some environmental issues like sea levels, wind speeds and pollution concentrations, where there is a potential for catastrophic results but it happens rarely. Recently, EVT has been used increasingly in finance and insurance.

The main difficulty of estimating rare events is that in most cases there is a small amount of, or even no, data available. The EVT approach is to develop models based on asymptotic theory. EVT models the limiting distribution of the extreme values of a random variable, which corresponds to the happening of rare events. A description of the method is beyond the scope of this document, however, several useful references are cited in the bibliography.

Regression

Often it's necessary and useful to develop a model of a variable by examining its drivers or causal variables. A regression equation expresses a dependent variable as a function of one or more predictor variables. Regression equations provide managers more information on the dynamics underlying a specific risk to help manage, insure or hedge the risk.

Methods Based on a Combination of Historical Data and Expert Input

Often there is not enough data to reliably quantify risks directly through data analysis. In these cases it's necessary to develop a model of the underlying dynamics that give rise to the data. This requires drawing on the experience and knowledge of domain experts to fill in the data gaps. The following methods attempt to model the dynamics of a system by using a combination of both historical data and expert input.

System Dynamics Simulation

System Dynamics is a robust modeling method that explicitly simulates the cause/effect relationships underlying the dynamics of system. The approach leverages both existing historical data and the knowledge and experience of senior managers to develop a stochastic simulation model. The model is used to run Monte Carlo simulations and develop probability distributions for the variable of interest.

The System Dynamics approach has several advantages over parametric approaches described above, particularly for modeling operational risks:

- It provides a systematic way to fill any gaps in historical data with input from experts relying on their knowledge and experience. This is applicable particularly for modeling operational risks where it's often the case that there isn't enough representative data to apply the statistical methods described above.
- It provides a way to determine how operational risks change as a function of changes in operations. Since the approach explicitly captures the cause/effect linkages, it is easier to develop effective ways to mitigate risk and measure their impact than with noncausal methods.
- As businesses become more complex, knowledge of their underlying dynamics becomes more fragmented and localized. Although many managers have a good understanding of their own functional areas, few have a solid grasp of the dynamics of the entire organization. Obtaining a complete picture, for example, of the sources of operational risks and how they affect financial performance, requires the combined knowledge of managers across functional areas. The system dynamics approach facilitates this interaction through a structured, participative modeling and decision-making process.

Fuzzy Logic

In spite of its name, fuzzy logic is a well-established engineering science used successfully in control systems and expert reasoning. It is an approach to modeling complex systems, where much of the complexity comes from the ambiguous, uncertain or undecided representation of the variables of the system. Traditional quantitative models tend to interpret reality in binary terms. For example, imagine a device that identifies if a person has a fever. Given the temperature of an individual, a quantitative model programmed in the device will use a discrete, binary rule, such as: "if the temperature is at or over 103°F then person has a fever, else normal". Even if it has other categories in between, such as "light fever", it will still use a discrete binary rule to determine whether a person falls in the "light fever" category or "fever" category. However, in reality it's clear that there is no precise cut-off for determining whether someone has a fever and the boundary between "normal" and "fever" is fuzzy. Fuzzy set theory was developed to recognize these gray areas. According to fuzzy set theory, a person with a temperature of 101.5°F would be classified as having some membership in both categories "normal" and "fever". Fuzzy logic is the reasoning based on fuzzy set theory.

Fuzzy logic has advantages in modeling complex business problems where linguistic variables are used to express the logic rules, the information is subjective, incomplete or unreliable, and the problem spaces are often nonlinear. A fuzzy system is closer to the way people reason and is therefore often used to build expert systems. The fuzzy nature of the rule spaces makes it easy to model multiple, often different or conflicting expert views toward the same model variables. In terms of risk modeling and assessment, fuzzy logic shows potential to be a good approach in dealing with operational risk, where the probability assessment is often based on expert opinion and the risk space is multidimensional and highly nonlinear.

Estimating Probabilities through Expert Testimony

In extreme cases, there aren't any data at all. In these cases, one must rely on the knowledge and experience of domain experts. Probability distributions for events for which there is sparse data can be estimated through expert testimony. A naive method for assessing probabilities is to ask the expert, e.g., "What is the probability that a new competitor will enter the market?" However, the expert may have difficulty answering direct questions and the answers may not be reliable. Behavioral scientists have learned from extensive research that the naive method can produce unreliable results because of heuristics and biases. For example, individuals tend to estimate higher probabilities for events that can be easily recalled or imagined. Individuals also tend to anchor their assessments on some obvious or convenient number resulting in distributions that are too narrow. (See Clemen, 1996 and von Winterfeldt & Edwards, 1986 in the bibliography for further examples). Decision and risk analysts have developed several methods for accounting for these biases. Several of these methods are described below.

Preference among Bets

Probabilities are determined by asking the expert to choose which side she prefers on a bet on the underlying events. To avoid issues of risk aversion, the amounts wagered should not be too large. For example, a choice is offered between the following bet and its opposite:

Bet	Opposite Side of Bet
Win $\$x$ if a new competitor enters the market Lose $\$y$ if no new competition	Lose $\$x$ if a new competitor enters the market Win $\$y$ if no new competition.

The payoffs for the bet, amounts $\$x$ and $\$y$, are adjusted until the expert is indifferent to taking a position on either side of the bet. At this point, the expected values for each side of the bet are equal in her mind. Therefore,

$$\$xP(C) - \$y[1-P(C)] = -\$xP(C) + \$y[1-P(C)]$$

where $P(C)$ is the probability of a new competitor entering the market. Solving this equality for $P(C)$:

$$P(C) = \$y / (\$x + \$y)$$

For example, if the expert is indifferent to taking a position on either side of the following bet:

Win $900 if a competitor enters the market
Lose $100 if no new competition

then the estimated subjective probability of a new competitor entering the market is $100/($100 + $900) = 0.10.

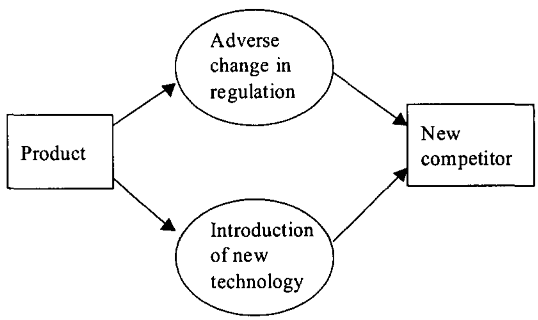
Judgments of Relative Likelihood

This method involves asking the expert to provide information on the likelihood of an event relative to a reference lottery. The expert is asked to indicate whether the probability of the event occurring is more likely, less likely or equally likely compared to a lottery with known probabilities. Typically a spinning wheel (a software implementation of the betting wheels in casinos) is used on which a portion of the wheel is colored to represent the event occurring. The relative size of the colored portion is specified. The expert is asked to indicate whether the event is more, less or equally likely to occur than the pointer landing on the colored area if the wheel was spun fairly. The colored area is reduced or increased as necessary depending on the answers until the expert indicates that the two events are equally likely. This method is often used with subjects that are naive about probability assessments.

Decomposition to Aid Probability Assessment

Often, decomposing an event into conditional causal events helps experts assess risk of complex systems. The structure of the conditional causal events can be represented by an **influence diagram.** Influence diagrams illustrate the interdependencies between known events (inputs), scenarios and uncertainties (intermediate variables) and an event of interest (output). An influence diagram model comprises risk nodes representing the uncertain conditions surrounding an event or outcome. Relationships among nodes are indicated by connecting arrows, referred to as arcs of influence. The graphical display of risks and their relationships to process components and outcomes helps users visualize the impacts of external uncertainties.

While this approach increases the number of probability assessments, it also allows input from multiple experts or specialists, and helps combine empirical data with subjective data. For example, a new competitor entering the market may be decomposed using an influence diagram such as this one:



The probability of a new competitor, P(C) can be estimated, using a **Bayesian approach**. The approach uses "Bayes' Rule" which is a formal, optimal equation for the revision of probabilities in light of new evidence contained in conditional or causal probabilities.

$$P(C) = \sum_i P(C_i \mid R_i, T_i)\, P(R_i, T_i)$$

where i is a product index, $P(R_i, T_i)$ is the joint probability of an adverse change in regulation and introduction of new technology, and $P(C_i \mid R_i, T_i)$ is the conditional probability of a new competitor entering a market for product i. This formula is useful when assessing the conditional probabilities $P(C_i \mid R_i, T_i)$ and is easier than a direct calculation of P(C).

Several different experts may be asked to assess the conditional and joint probabilities. For example, one expert (or group of experts) may assess the probability of adverse regulation for a specific product, another expert may assess probability of introduction of new technology and a third may assess the probability of a new competitor given the state of new regulation and technology.

The Delphi Technique

Scientists at the Rand Institute developed the "Delphi process" in the 1950's for forecasting future military scenarios. Since then it has been used as a generic strategy for developing consensus and making group decisions, and can be used to assess probabilities from a group of individuals. This process structures group communication, and usually involves anonymity of responses, feedback to the group as collective views, and the opportunity for any respondent to modify an earlier judgment. The Delphi process leader poses a series of questions to a group; the answers are tabulated and the results are used to form the basis for the next round. Through several iterations, the process synthesizes the responses, resulting in a consensus that reflects the participants' combined intuition, experience and expert knowledge.

The Delphi technique can be used to explore or expose underlying assumptions or information leading to differing judgments and to synthesize informed judgments on a topic spanning a wide range of disciplines. It is useful for problems that can benefit from subjective judgments on a collective basis.

Pitfalls and Biases

Estimating subjective probabilities is never as straightforward as implied in the description of the methods above. There are several pitfalls and biases to be aware of:

None of the methods works extremely well by itself. Typically, multiple techniques must be used. To increase consistency, experts should be asked to assess both the probability of an event and, separately, the probability of the complement of the event. The two should always add up to 1.0; however, in practice they seldom do without repeated application of the assessment method. The events must be defined clearly to eliminate ambiguity. "What is the probability of a new competitor entering the market?" is an ambiguous question. "What is the probability that a new competitor will take more than 5% market share of product A in the next two years?" is much less ambiguous and more clearly defines the event. When assessing probabilities for rare events, it is generally better to assess odds. Odds of event E is [P(E) / P(complement of E)].

* * * * *

Note: This appendix was reproduced from the Tillinghast – Towers Perrin monograph *RiskValueInsights®: Creating Value Through Enterprise Risk Management*, (http://www.tillinghast.com).

Appendix C — ERM Bibliography

Books

Arthus, Mark G., *Integrated Compliance and Total Risk Management: Creating a Bankwide Compliance System That Works*, (McGraw-Hill, March 1994)

Bernstein, P.L., *Against the Gods: The Remarkable Story of Risk*, (John Wiley & Sons, 1996)

Borodovsky, Lev, *Practitioner's Handbook of Financial Risk Management*, (Butterworth-Heinemann, May 2000)

Cannon, Tom, *A Guide to Integrated Risk Management*, (London: AIRMIC, 1999)

Clemen, Robert T., *Making Hard Decisions*, 2nd edition, (Duxbury Press, 1996)

Cox, J.J. and Tait, N., *Reliability, Safety and Risk Management*, (Butterworth-Heinemann, January 1991)

DeLoach, James, *Enterprise-Wide Risk Management*, (FT Prentice Hall, 1998)

Doherty, Neil A., *Integrated Risk Management: Techniques and Strategies for Managing Corporate Risk*, (McGraw Hill, 2000)

Dowd, Kevin, *Beyond Value at Risk: The New Science of Risk Management*, (John Wiley & Sons, Inc., February 1998)

Ehrbar, Al, *EVA: The Real Key to Creating Wealth*, (Stern Stewart & Co., 1998)

Fusaro, Peter C., *Energy Risk Management: Hedging Strategies and Instruments for the International Energy Markets*, (McGraw-Hill, April 2, 1998)

Gastineau, Gary L. and Kritzman, Mark P., *Dictionary of Financial Risk Management*, (Irwin Professional Pub., January 15, 1996)

Grant, James and Fabozzi, Frank J., *Foundations of Economic Value Added*, (1997)

Gruening, Hennie Van, *Analyzing Banking Risk: A Framework for Assessing Corporate Governance and Financial Risk Management*, (World Bank, April 1999)

Haimes, Yacov Y., *Risk Modeling, Assessment, and Management*, (Wiley-Interscience, August 1998)

Hull, J.C., *Options, Futures and Other Derivatives*, (Prentice-Hall, 2000)

Overview of Enterprise Risk Management

Jorion, Phillipe, *FRM Handbook*, (Wiley/GARP Risk Management Library, 2001)

Kendall, Robin, *Risk Management For Executives*, (FT Prentice Hall, 1998)

McNamee, David, *Business Risk Assessment*, (Altamonte Springs, FL, Institute of Internal Auditors Research Foundation, 1998)

McNamee, David and Georges, Selim, *Risk Management: Changing The Internal Auditor's Paradigm*, (Altamonte Springs, FL, Institute of Internal Auditors Research Foundation, 1998)

Melchers, R.E., *Integrated Risk Assessment: Applications and Regulations*, (Newcastle, Australia, May 1998)

Panjer, Harry H. et al, *Financial Economics: With Applications to Investments, Insurance and Pensions*, (The Actuarial Foundation, Schaumburg, IL, 1998)

PricewaterhouseCoopers, *The Regulatory Risk Management Handbook: 1998-1999*, (Sharpe, M.e., Inc., September 1998)

Smithson, Charles W. and Clifford W. Smith, *Managing Financial Risk: A Guide to Derivative Products, Financial Engineering and Value* (McGraw Hill, 3rd edition, July 1998)

Stern, Joel M., Shiely, John S., et. al., *The EVA Challenge*

von Winterfedt, D. and W. Edwards, *Decision Analysis and Behavioral Research*, (Cambridge University Press, 1986)

White, Leslie and Herman, Melanie, *Leaving Nothing to Chance: Achieving Board Accountability through Risk Management*, (National Center for Nonprofit Boards, October 1998)

Young, Peter C. and Fone, Martin, *Organization Risk Management in UK Police Authorities: An Integrated Management Approach*, (*White Paper. University of St. Thomas*)

Young, Peter C. and Tippins, Steven C., *Managing Business Risk: An Organization-Wide Approach to Risk Management*, (AMACOM, September 2000)

Zask, Ezra, *Global Investment Risk Management*, (McGraw-Hill, September 1999)

Ziemba, William T. and Mulvey, John M., *Worldwide Asset and Liability Modeling*, (Cambridge University Press, 1998)

Overview of Enterprise Risk Management

<u>Articles</u>

A.M. Best

"A New World of Risks," *Best's Review Life/Health Insurance Edition* (January 2000, pg. 3)

"A.M. Best Enterprise Risk Model – A Holistic Approach to Measuring Capital Adequacy," (July 2001)

Green, Meg, "Microscope Or Binoculars," *Best's Review (April 2003, pg. 21)*

McDonald, Lee, "Ryan's Hope: Change and More Change," *Best's Review Life/Health Insurance Edition* (March 1998, pg. 84)

Miccolis, Jerry A., "All Together Now," *Best's Review* (February 2000, pg. 122)

Whitney, Sally, "Managing Internal Risks," *Best's Review* (April 2000, pg. 141)

Business Insurance

"Expanding the Envelope," *Business Insurance* (April 12, 1999, pg. 116)

"New Chief Risk Officer Role Coordinates Risk Strategy," *Business Insurance* (April 26, 1999, pg. 3)

"Integrated Products Support Expanded Risk Manager Duties," *Business Insurance* (April 26, 1999, pg. 3)

"Balance Sheet Risks," *Business Insurance* (April 26, 1999, pg. 36)

"Enterprise Risk Management Utilized: Competitive Market Means Power Companies Must Consider New Exposures," *Business Insurance* (August 2, 1999, pg. 35)

"A Risk Approach For The Enterprising," *Business Insurance* (October 18, 1999, pg. 61)

"Enterprise, Integrated Programs Expand Choices," *Business Insurance* (November 15, 1999, pg. 50)

"Enterprise Risk Policy Crafted," *Business Insurance* (January 31, 2000, pg. 2)

"Top Priority on Bottom Line," *Business Insurance* (March 20, 2000, pg. 3)

"Insurers Stymied in Managing Enterprise Risk," *Business Insurance* (April 17, 2000, pg. 1)

Overview of Enterprise Risk Management

"Spencer Grants $300,000," *Business Insurance* (May 15, 2000, pg. 1)

"CRO Role Suited to Enterprise Risk Plans," *Business Insurance* (May 15, 2000, pg. 16)

"'Holistic' View Debated: Few Products For Transferring Broad Array of Risk," *Business Insurance* (June 5, 2000, pg. 33)

"Enterprise Risk Management: Number of Chief Risk Officers Grows as Concept Takes Hold," *Business Insurance* (September 25, 2000, pg. 3)

"Managing Risk Adds Value," *Business Insurance* (October 30, 2000, pg. 2)

"Definition Varies, But Enterprise Risk Cover Draws Interest," *Business Insurance* (November 13, 2000, pg. 3)

"ERM RMIS Requires Corporate Cultural Transformation," *Business Insurance* (December 4, 2000, pg. 21).

"Interest Wanes in Earnings Cover: Policy Takes Downhill Turn," *Business Insurance* (December 11, 2000, pg. 1).

"Public Entity Adopts Holistic Plan: Enterprising Risk Manager," *Business Insurance* (January 1, 2001, pg. 1)

"Human Capital Also an Enterprise Risk," *Business Insurance* (January 1, 2001, pg. 8)

Financial Times

Brown, Gregory W., "Mastering Risk: Seeking Security in a Volatile World," *Financial Times* (May 16, 2000)

Butterworth, Mark, "Mastering Risk: The Emerging Role of the Risk Manager," *Financial Times* (April 25, 2000)

Dickinson, Gerry, "Mastering Management: Risk Role Grows to Enterprise Scale," *Financial Times* (November 13, 2000)

Finger, Christopher C. and Malz, Allan M., "Mastering Risk: Welcome to This Week's One in a Million Event," *Financial Times* (June 6, 2000)

Glasserman, Paul, "Mastering Risk: The Quest for Precision Through Value-at-Risk," *Financial Times* (May 16, 2000)

Hanley, Mike, "Mastering Risk: Lowering Exposure by Spreading the Risk," *Financial Times* (May 2, 2000)

Overview of Enterprise Risk Management

Jorion, Philippe, "Mastering Risk: Value, Risk and Control: A Dynamic Process in Need of Integration," *Financial Times* (May 16, 2000)

Knight, Rory and Pretty, Deborah, "Mastering Risk: Philosophies of Risk, Shareholder Value and the CEO," *Financial Times* (June 27, 2000)

Meulbroek, Lisa, "Mastering Risk: Total Strategies for Company-Wide Risk Control," *Financial Times* (May 9, 2000)

Mitroff, Ian, "The Essentials of Crisis Management," *Financial Times* (June 20, 2000)

Stulz, Rene, "Mastering Risk: Diminishing the Threats to Shareholder Wealth," *Financial Times* (April 25, 2000)

Financing Risk & Reinsurance

Clark, Brent, "Enterprise Risk - - What's Up With That?," *Financing Risk & Reinsurance* (October 2000)

Miccolis, Jerry A., "Enterprise Risk Management: What's Beyond The Talk?," *Financing Risk & Reinsurance* (May 2000)

Miccolis, Jerry A., "Enterprise Risk Management in the Financial Services Industry: Still a Long Way To Go," *Financing Risk & Reinsurance* (August 2000)

Miccolis, Jerry A., "Enterprise Risk Management in the Financial Services Industry: From Concept to Management Process," *Financing Risk & Reinsurance* (November 2000)

Thomas, Bruce B., "Evaluating Catastrophe Indices," *Financing Risk & Reinsurance* (January 1998)

Thomas, Bruce B., "Index-Linked Hedges," *Financing Risk & Reinsurance* (September 1998)

Thomas, Bruce B., "Mitigating Basis Risk in Index-Linked Securities," *Financing Risk & Reinsurance* (October 1998)

Thomas, Bruce B., "Index-Based Reinsurance," *Financing Risk & Reinsurance* (January 1999)

Thomas, Bruce B., "Hedging Peak Loss Experience," *Financing Risk & Reinsurance* (February 1999)

Thomas, Bruce B., "Basis Opportunity," *Financing Risk & Reinsurance* (June 1999)

Overview of Enterprise Risk Management

Thomas, Bruce B., "The Road Not Taken," *Financing Risk & Reinsurance* (September 1999)

Thomas, Bruce B., "Measuring Up: Benchmarking Underwriting and Loss Performance," *Financing Risk & Reinsurance* (January 2000).

Moody's Investors Service

"One Step in the Right Direction: The New C-3a Risk-Based Capital Component" (June 2000)

National Underwriter

"Firm Touts 'Enterprise Risk Management,'" *National Underwriter/Property & Casualty Edition* (June 23, 1997, pg. 19)

"Business-Risk Talk Spawns Little Action," *National Underwriter/Property & Casualty Edition* (May 11, 1998, pg. 25)

"Risk Mgt. Consultants Focus on Offering Top-Down Programs," *National Underwriter/Property & Casualty Edition* (June 8, 1998, pg. 3)

"'Enterprise Risk Management' Fuels Captive Surge," *National Underwriter/Property & Casualty Edition* (October 26, 1998, pg. S21)

"Integrating Financial, Insurance Risks Takes Team Effort," *National Underwriter/Property & Casualty Edition* (October 26, 1998, pg. S32)

"CRO Career Dream Alive Among RMs," *National Underwriter/Property & Casualty Edition* (May 17, 1999, pg. 25)

"RM Successes Called Career Launch Pad," *National Underwriter/Property & Casualty Edition* (February 14, 2000, pg. 19)

"M&M Unit Tackles Enterprise Risk," *National Underwriter/Property & Casualty Edition* (April 3, 2000, pg. 17)

"Insurers Seen Stalling Enterprise Risk Management," *National Underwriter/Property & Casualty Edition* (April 10, 2000, pg. 9)

"Insurance Executives Doubt Their RM Abilities," *National Underwriter/Property & Casualty Edition* (May 29, 2000, pg. 15)

"Enterprise RM Must Start With The Data," *National Underwriter/Property & Casualty Edition* (June 5, 2000, pg. 12)

Overview of Enterprise Risk Management

"How Much of 'Operational' Risk Management is Hype?," *National Underwriter/Property & Casualty Edition* (June 5, 2000, pg. 15)

Risk Management

Berry, Andrew and Phillips, Julian, "Enterprise Risk Management: Pulling It Together," *Risk Management* (September 1998, pg. 53)

Berry, Andrew, "Future Shock? An Industry Forecast," *Risk Management* (April 2000, pg. 25)

Busman, Evan R. and Van Zuiden, Paul, "The Challenge Ahead: Adopting an Enterprise-wide Approach to Risk," *Risk Management* (January 1998, pg. 14)

Conley, John, "Multiple Lines: A Status Report on The Building Trend," *Risk Management* (January 2000, pg. 19)

Conley, John, "Waves of the Future," *Risk Management* (July 1999, pg. 13)

Conley, John, "3 Winning Ways," *Risk Management (*December 1999, pg. 12)

Eiss, Elizabeth, "Enterprising Solutions," *Risk Management* (August 1999, pg. 34)

Hereth, Mark L., "Beyond The Box," *Risk Management* (March 1996, pg. 29)

Hernandez, Luis Ramiro, "Integrated Risk Management in The Internet Age," *Risk Management* (June 2000, pg. 29)

Herrick, R.C., "Exploring The Efficient Frontier," *Risk Management* (August 1997, pg. 23)

Lam, James C. and Brian M. Kawamoto, "Emergence of The Chief Risk Officer," *Risk Management* (September 1997, pg. 30)

Lange, Scott, "Going 'Full Bandwidth' at Microsoft," *Risk Management* (July 1996, pg. 29)

Lee, Charles R., "Chief Risk Officer Stepping Up," *Risk Management* (September 2000, pg. 22)

Levin, Michael R. and Rubenstein, Michael, L., "A Unique Balance: The Essence of Risk Management," *Risk Management* (September 1997, pg. 37)

Overview of Enterprise Risk Management

Miccolis, Jerry A., "Toward a Universal Language of Risk," *Risk Management* (July 1996, pg. 45)

Miccolis, Jerry A. and Quinn, Timothy P., "What's Your Appetite For Risk? Determining The Optimal Retention," *Risk Management* (April 1996, pg. 41)

Pelland, Dave, "Greater Emphasis on Financial Skills: Changing Face of Risk Management," *Risk Management* (April 1997, pg. 108)

Rahardjo, Kay and Dowling, Mary Ann, "A Broader Vision: Strategic Risk Management, "*Risk Management* (September 1998, pg. 44)

Sanderson, Scott M. and Koritzinsky, Arthur G. "Risk Centers: Enterprise-Wide Efficiency," *Risk Management (*March 1999, pg. S6)

Thomas, Bruce B., "Insurance Meets Wall Street," *Risk Management* (May 1997)

Thornhill, Wil and Derksen, Amy, "A United Approach: Creating Integrated Risk Plans," *Risk Management* (August 1998, pg. 36)

Williams, Todd L., "An Integrated Approach to Risk Management," *Risk Management* (July 1996, pg. 22)

Williams, Todd L., "Convergence," *Risk Management* (August 1999, pg. 13)

Zarb, Frank G., "Brokers at The Helm: Navigating The Risk Financing Frontier," *Risk Management* (July 1995, pg. 53)

Standard and Poor's

"Analyzing Capital Adequacy for Property/Casualty Insurers," (March 1999)

"Property/Casualty Insurance Ratings Criteria"

"Revised Risk-Based Capital Adequacy Model for Financial Products Companies"

Treasury & Risk Management and CFO Magazines

"Treasury in an Uncertain World," *Treasury & Risk Management* (March 1999)

Banham, Russ, "Innovative Risk-Transfer Packages Aren't Making The Cut," *Treasury & Risk Management* (October 1999, pg. 39)

Banham, Russ, "Kit and Caboodle: Understanding The Skepticism About Enterprise Risk Management," *CFO Magazine* (April 1999, pg. 63)

Overview of Enterprise Risk Management

Banham, Russ, "Top Cops of Risk," *CFO Magazine* (September 2000, pg. 91)

Banham, Russ, "Whatever The Weather: How United Grain Growers Tamed Mother Nature in Completing The Deal of The Decade," *CFO Magazine* (June 2000, pg. 117)

Ewing, Lace, "Don't Buy Anything Until You Talk to Your Risk Manager," *Treasury & Risk Management* (May/June 2000, pg. 5)

Hulihan, Maile, "Enterprise-Wide Needs," *Treasury & Risk Management* (July 1999)

Moules, Jonathan, "The Next Frontier," *Treasury & Risk Management* (July 1999)

Price, Margaret, "Under Control," *Treasury & Risk Management* (October 1998, pg. 30)

Seuntjens, Thomas, "Look, Boss I Cut the Premiums!," *Treasury & Risk Management* (January/February 2000, pg. 8)

Teach, Edward, "Microsoft's Universe of Risk," *CFO Magazine* (March 1997)

Wall Street & Technology

"Enterprise Risk Management Solution," *Wall Street & Technology* (October 1996, pg. 81)

"Hedging Credit, Market Risk," *Wall Street & Technology* (March 1998), pg. 23

"The New Risk Management," *Wall Street & Technology* (July 1998), pg. 88

"ERM a Priority Post Y2K, Spending on The Rise," *Wall Street & Technology* (July 1999, pg. 114)

"Risk Management Tech Spending on the Rise," *Wall Street & Technology* (October 1999, pg. 7)

"Pulling It All Together," *Wall Street & Technology* (October 1999, pg. 7)

"Risk Vendors Jump on The Buy-side Bandwagon," *Wall Street & Technology* (October 1999, pg. 16)

"Crystal Box Offers Clear View of Risk Management at CDC," *Wall Street & Technology* (March 2000, pg. 64)

"Risk: Enterprise-wide Risk Technology Spending Reaches a Plateau, Credit Still Rising," *Wall Street & Technology* (October 2000, pg. 54)

Overview of Enterprise Risk Management

Wall Street Journal

Dvorak, Phred, "Japan's Largest Insurer Loses $2 Billion in Gamble on Euro-Denominated Bonds," *The Wall Street Journal* (June 5, 2000, pg. A22)

Kranhold, Kathryn and Erin White, "The Perils And Potential Rewards of Crisis Managing For Firestone," *The Wall Street Journal* (September 8, 2000, pg. B1)

Other Publications

"A Generation of Risk Managers Fortell a Future," *ABA Banking Journal* (October 1998, pg. 74)

"Risk Becomes More Enterprising," *Airfinance Journal* (December 1998, pg. 28)

"Insurers Put Pressure on Capital Markets," *Airfinance Journal* (September 1999, pg. 22)

"Ex-Bankers Trust President Starts Risk Management Firm," *American Banker* (June 18, 1997, pg. 18)

"Financial Management, Saving Dollars by Managing Risk," *American City & County* (1998)

"Safe and Sound," *American Gas* (September 1999, pg. 24)

"Modeling and Comparing Dependencies in Multivariate Risk Portfolios," *ASTIN Bulletin* (1998)

"Foretelling The Future," *Australian CPA* (August 2000, pg. 44)

"Choosing an ERM System That Suits Your Bank's Needs," *Bank Accounting & Finance* (Fall 1999, pg. 39)

"Competitive Strategies For a New Era," *Bank Systems & Technology* (August 1996, pg. 48)

"Risk Management: Banks Wager on Enterprise Systems," *Bank Systems & Technology* (September 1998, pg. 30)

"Technotes," *Banker* (March 1998, pg. 74)

"A New Role: Chief Risk Officer," *Business Journal* (November/December 1999, pg. 55)

"Holistic Risk on The Horizon," *Canadian Underwriter* (June 1999, ɔg. 12)

Overview of Enterprise Risk Management

"Integrated Risk Programs Gain Ground," *Canadian Underwriter* (April 2000, pg. 22)

"Insurance Executives Need Better Enterprise Risk Management," *Community Banker* (May 2000, pg. 42)

"Systems Deliver Functionality But Falter on Integration," *Computerworld* (February 10, 1997, pg. 57)

"Insurance Broker Predicts Utility Trends," *Electric Light & Power* (June 1997, pg. 8)

"Risk Managing Shareholder Value," *Emphasis* (1999/3, pg. 18)

"Risk Management: Getting a Handle on Operational Risks," *Emphasis* (2000/1, pg. 22)

"The Three P's of Total Risk Management," *Financial Analysts Journal* (January/February 1999, pg. 13)

"What Risk Premium is Normal," *Financial Analysts Journal* (2002)

"Risky Business," *Financial Executive* (September/October 1998, pg. 30)

"Enterprise Risk Management: What's Driving The Trend?," *Financial Executive* (September/October 1998, pg. 32)

"Earnings Insurance: Mission Impossible?," *Financial Executive* (May/June 1999, pg. 53)

"Chief Risk Officers - How Will Their New Role Change Finance?," *GARP Newsletter* (December 2000)

"New Job Title: Chief Risk Officer," *Global Finance* (March 1998)

"A Better Way to Manage Risk," *Harvard Business Review* (February 2001)

"A Framework for Risk Management," *Harvard Business Review* (November/December 1994)

"Scientific Management at Merck: An Interview With CRO Judy Lewent," *Harvard Business Review* (January-February 1994)

"Honeywell, Inc. and Integrated Risk Management," *Harvard Business School Case* (July 12, 2000)

"Integrated Risk Management for the Firm: A Senior Manager's Guide," *Harvard Business School's Division of Research Working Papers* (2000-2001)

"Harvard Business Review on Managing Uncertainty," *Harvard Business School Press* (January 1999)

Overview of Enterprise Risk Management

"Managing Risk - All On One Page," *Harvard Management Update* (Nov 1998, pg. 12)

"Risky Business," *Industry Week* (May 15, 2000, pg. 75)

"Survey Of Risk Managers Reveals Dissatisfaction With Prioritizing Techniques," *Insurance Advocate* (May 2000, pg. 18)

"A Synthesis of Risk Measures," *Insurance Mathematics and Economics* (1999)

"On Two Dependent Individual Risk Models," *Insurance Mathematics and Economics* (2001)

"On the Distribution of a Sum of Correlated Aggregate Claims," *Insurance Mathematics and Economics* (1998)

"On the Distribution of Two Classes of Correlated Aggregate Claims," *Insurance Mathematics and Economics* (1999)

"The Discrete-Time Risk Model With Correlated Classes of Business," *Insurance Mathematics and Economics* (2000)

"The Impact of Dependence Among Multiple Claims in a Single Loss," *Insurance Mathematics and Economics* (2000)

"Street Fighters," *International Business* (September/October 1997, pg. 19)

"People Issues Are Firms' Top Risk," *Journal of Commerce* (May 30, 2000, pg. 8)

"Value at Risk Calculations, Extreme Events and Tail Estimation," *Journal of Derivatives* (Spring 2000)

"Enterprise-wide Risk Management: Staying Ahead of the Convergence Curve," *Journal of Lending & Credit Risk Management* (June 1999, pg. 16)

"Enterprise-wide Risk Management," *Journal of Lending & Credit Risk Management* (March 2000, pg. 22)

"Incorporating Volatility Updating Into the Historical Simulation Method for Value-at-Risk," Journal of Risk (1998)

"Capturing the Correlations of Fixed-Income Instruments," *Management Science* (October 1994, pg. 1329)

"Coherent Measures of Risk," *Mathematical Finance* (1999, pg. 203)

Overview of Enterprise Risk Management

"Risk Management and Analysis," *Measuring and Modeling Financial Risk* (March 1999)

"Countering Strategic Risk With Pattern Thinking," *Mercer Management Journal* (November 1999, pg. 11)

"Managing Risk Through an Enterprise Approach," *Mercer on Transport* (Fall/Winter 1999, pg. 29)

"More Companies Embrace Enterprise Risk Management," *Office.com* (May 10, 2000)

"Risk Management: Where Utilities Still Fear to Tread," *Public Utilities Fortnightly* (October 15, 2000, pg. 40)

"The Final Frontier of Risk," *ReActions* (May 1999, pg. 20)

"An Enterprising View of Risk," *ReActions* (May 1999, pg. 5)

"Correlation: Pitfalls and Alternatives," *Risk* (1999)

"Risk and Insurance: The Hidden Links," *Risk* (January 2001, pg. 54)

"Integrated Risk Assessment: Current Views of Risk Management," *Risk Management Reports* (April 1998, pg. 2)

"Integrating Risk Management," *Risk Management Reports* (April 1997, pg. 2)

"Internal Auditors," *Risk Management Reports* (November 1999, pg. 2)

"Enterprise Risk: All Wrapped Up To Go," *Risk & Insurance* (October 1, 1999, pg. 1)

"The Ever-Widening Spectrum of Risk," *Risk and Insurance* (October 1, 2000, pg. 43)

"The Coming Revolution in Risk Management," *Risk and Rewards* (October 2002)

"Global Risk Management," *Strategic Finance* (August 1999)

"Enterprise Risk Management," *Strategy & Leadership* (March/April 1998, pg. 10)

"Saving Dollars by Managing Risk," *The American City & County* (April 1998, pg. 8)

"Not Too Risky a Business," *The Banker* (July 1997, pg. 90)

"Buying a Financial Umbrella," *The Economist* (June 2000, pg. 75)

"Expanding Risks: Enterprise Risk Management Solutions," *The John Liner Review* (Spring 2000, pg. 63)

"Is Your Enterprise in Good Hands?," *Transportation & Distribution* (June 1999, pg. SCF2)

"Countering The Domino Effect," *U.S. Banker* (August 1999, pg. 44)

"Homogenizing Catastrophe Risk," *Viewpoint* (Fall 1997)

"Operating Risk Goes Systematic," *Viewpoint* (Winter 1998-1999)

<u>Service Provider Publications</u>

Aon Risk Services

"Enterprise Risk Management: Part One," *Aon Insights* (Edition 3, 1999)

"Enterprise Risk Management: Part Two," *Aon Insights* (Edition 4, 1999)

"Investigating Enterprise Risk Management: Part One," *Aon Insights* (Edition 1, 2000)

"Investigating Enterprise Risk Management: Part Two," *Aon Insights* (Edition 2, 2000)

"Investigating Enterprise Risk Management: Part Three," *Aon Insights* (Edition 4, 2000)

"Enterprise Risk Management," <http://www.aon.com/solutions/prod_serv/ps_28.asp>

"What is ERM?," <http://www.aon.com/solutions/prod_serv/ERM/erm1.asp>

"ERM Trends," <http://www.aon.com/solutions/prod_serv/ERM/erm2.asp>

"ERM FAQ's: What's on The CEO's Mind When it Comes to Risk?," <http://www.aon.com/solutions/prod_serv/ERM/erm3.asp>

"ERM Solutions," <http://www.aon.com/solutions/prod_serv/ERM/erm4.asp>

"ERM Expert Profile: Another Perspective on Corporate Governance," <http://www.aon.com/solutions/prod_serv/ERM/erm9.asp>

Punter, Alan, "The Changing Economics of Non-Life Insurance: New Solutions for the Financing of Risk," *White Paper, Aon Group Limited* (March 29, 2000, 23 pgs)

eRisk

Khwaja, Amir, "Enterprise-wide Risk Management and the Impact of XML," *eRisk* (February 17, 2000)

Overview of Enterprise Risk Management

Lam, James C., "Enterprise-wide Risk Management and the Role of the Chief Risk Officer," *eRisk* (March 25, 2000)

Tillinghast – Towers Perrin

"Enterprise Risk Management in The Insurance Industry: 2000 Benchmarking Survey Report," (2000) (http://www.tillinghast.com)

"Enterprise Risk Management in The Insurance Industry: 2002 Benchmarking Survey Report," (2002) (http://www.tillinghast.com)

Miccolis, Jerry A. and Shah, Samir, "Enterprise Risk Management: An Analytic Approach," *Tillinghast – Towers Perrin Monograph* (January 2000) (http://www.tillinghast.com)

Miccolis, Jerry A. and Shah, Samir, "RiskValueInsights™: Creating Value Through Enterprise Risk Management – A Practical Approach for the Insurance Industry," *Tillinghast – Towers Perrin Monograph* (April 2001) (http://www.tillinghast.com)

Other

Conference Board of Canada

Birkbeck, Kimberly, "Realizing The Rewards of Risk: How Integrated Risk Management Can Help Your Organization," *Proceedings of The 1997 International Conference on Risk Management* (April 1998)

Birkbeck, Kimberly, "Forewarned is Forearmed: Identification and Measurement in Integrated Risk Management," *Proceedings of The 1998 International Conference on Risk Management* (February 1999)

Birkbeck, Kimberly, "Integrating Risk Management: Strategically Galvanizing Resources in The Organization," *Proceedings of The 1997 International Conference on Risk Management* (April 1998)

Brown, Debra L. and Brown, David A.H., "Strategic Leadership for Effective Corporate Communications" *Proceedings of The 1999 International Conference on Risk Management* (February 2000)

Nottingham, Lucy, "A Conceptual Framework for Integrated Risk Management," *Conference Board of Canada* (September 1997)

Overview of Enterprise Risk Management

International Risk Management Institute

Miccolis, Jerry A., "The Language of Enterprise Risk Management: A Practical Glossary and Discussion of Relevant Terms, Concepts, Models and Measures" (http://www.irmi.com/expert/risk.asp)

Treasury Board of Canada Secretariat

"Best Practices in Risk Management: Private and Public Sectors Internationally," Treasury Board Secretariat Ottawa, Ontario, April 27, 1999

Overview of Enterprise Risk Management

The 2002 Enterprise Risk Management Committee

- Jerry Miccolis, chairman
- Paul Brehm
- Kevin Dickson
- Barry Franklin
- Gerry Kirschner
- John Kollar
- Don Mango
- François Morin
- Chris Nelson
- Ted Zubulake

Financial Pricing Models for Property-Casualty Insurance Products: Income Recognition and Performance Measurement

Sholom Feldblum, FCAS, FSA, MAAA, and
Neeza Thandi, FCAS, MAAA

Financial Pricing Models for Property-Casualty Insurance Products: Income Recognition and Performance Measurement

by Sholom Feldblum and Neeza Thandi

A financial pricing model determines premium rates that provide an adequate return on invested capital. The pricing of the policy depends on cash flows and capital requirements, both of which are external (exogenous) constraints. They do not depend on the accounting system used for performance measurement. Prospective pricing is not necessarily concerned with the *pattern* of profit recognition, which is an internal accounting construct.

Management compensation systems depend on the profit recognition pattern. We examine six accounting systems in this paper: statutory accounting, GAAP, tax accounting, fair value accounting, an NPV accounting system, and an IRR accounting system.

The profit recognition system is particularly important when managers change positions. Managers should be evaluated for their contributions to company performance. They should not be rewarded or penalized for the activities of past managers which affect the recognition of profits in later accounting periods.

Some of the most common performance measures, such as net after-tax operating income and the operating ratio, do not take account of the cost of capital. Both of these measures lead to perverse capital management incentives, since they reward managers for holding unnecessary capital.

We examine two measures of profitability for each accounting system discussed in this paper: (i) the rate of return and (ii) the economic value added.

1. The rate of return in each period is the net after-tax income divided by the beginning of the year capital.
2. The economic value added (EVA) in each period is the net after-tax income minus the dollar cost of using capital. The dollar cost of using capital is the product of the cost of capital and the amount of capital at the beginning of the year.[1]

IMPLIED EQUITY FLOW

For any accounting system, the implied equity flow equals the after-tax net income minus the change in capital. In this sentence, "capital" refers to the capital in the accounting system, not the invested capital.

Illustration: The implied equity flow equals the statutory income minus the change in required statutory surplus. The implied equity flow also equals the GAAP income minus the change in required GAAP equity.

---

1 Some analysts divide by the average capital held during the year, not the capital at the beginning of the year. This does not change the conclusions here, though the figures differ slightly.

The implied equity flow is the capital contribution required of the equityholders or the capital distribution to the equityholders. It is driven by statutory accounting. Once it is determined, it is a exogenous factor that is independent of the accounting assumptions used for performance measurement.

We illustrate the inter-relationships of the accounting systems with the following example.

Illustration: Statutory surplus is \$100 million at time t_i and \$105 million at time t_{i+1}. Net after-tax statutory income from time t_i to time t_{i+1} is \$10 million.

GAAP equity is \$125 million at time t_i, and GAAP after-tax income from time t_i to time t_{i+1} is \$10.5 million.

We solve for GAAP equity at time t_{i+1}. Let period t_{i+1} be the interval from time t_i to time t_{i+1}.

- Using the statutory figures, the implied equity flow during period t_{i+1} is \$10 million – (\$105 – \$100) = \$5 million.
- Using the GAAP figures, we have \$10.5 million – (\$X million – \$125 million) = \$5 million, or \$X million = \$130.5 million.

We compute the statutory return on surplus and the GAAP return on equity.

- The statutory return on surplus during period t_{i+1} is \$10 million ÷ \$100 million = 10%.
- The GAAP return on equity is \$10.5 million ÷ \$125 million = 8.4%.

We compare these figure with the cost of capital to determine the profitability of the insurance operations. Suppose that the cost of capital is 9% per annum.

- The statutory economic value added is \$10 million – 9% × \$100 million = +\$1 million. Using statutory accounting as the measure of profitability, the company has added \$1 million to equityholders' wealth.
- The GAAP economic value added is \$10.5 million – 9% × \$125 million = \$-0.75 million. Using GAAP accounting as the measure of profitability, the company has destroyed \$0.75 million of equityholders' wealth.

The implied equity flows are discussed in a companion paper (Feldblum and Thandi, [2002: Modeling the Equity Flows]). In a return on capital financial pricing model, the policy is priced by setting the net present value of the implied equity flows to zero or by setting the internal rate of return of the implied equity flows to the cost of equity capital.

The companion paper provides a fully documented illustration, showing all cash flows and implied equity flows for underwriting and investment operations. That illustration is continued in this paper, which deals with the income recognition pattern in various accounting systems.

Value of Operations

Neither GAAP nor statutory accounting is a proper reflection of the true value of the insurance operations. Statutory surplus and GAAP equity understate the net worth of the company because they do not recognize the capital embedded in the full value loss reserves. In addition, statutory accounting does not recognize a DPAC asset to offset the pre-paid acquisition costs in the gross unearned premium reserve.

For the same reasons, both statutory income and GAAP income may overstate or understate the change in the company's net worth during the accounting period. If the company is growing, statutory income and GAAP income understate the true change in the company's net worth. If the company is contracting, statutory income and GAAP income overstate the true change in the company's net worth.

We use an economic value added (EVA) yardstick to measure policy performance. Some readers may presume that the EVA measure corrects for the problems mentioned above. This is not the case. The EVA measure subtracts the cost of capital from the net income; it does not correct the distortions in the accounting system itself. If the net income measure is distorted, the EVA measure is distorted as well.

Net income and capital amounts depend on the accounting system. The list below summarizes the attributes shared by all the accounting systems discussed here.

- The company cash flows are the same for all accounting systems. These cash flows are exogenous; they are not affected by the performance measurement system.
- The implied equity flows are the same for all accounting systems. The implied equity flows are determined by the accounting system used for regulation; they are not affected by the accounting system used for performance measurement.
- The remaining capital after all losses are settled is zero for all accounting systems.
- The initial capital before policy inception is the initial implied equity flow. This is the same for all accounting systems.
- The sum of the nominal net income amounts (not the discounted net income amounts) is the same for all accounting systems. It equals the sum of the implied equity flows.

The final statement in the list above seems strange to some readers. They reason that if one accounting system recognizes income more rapidly than another accounting system, the investment income earned in the first accounting system is greater than the investment income earned in the second accounting system. This is not correct. The investment income is a cash flow, and the cash flows are invariant across accounting systems. The difference between accounting systems is that the investment income may be earned on assets backing required reserves in one accounting system and it may be earned on assets backing equity or net worth in another accounting system. Since the nominal income over the lifetime of the project equals the cash income over the lifetime of the project, the nominal income is invariant across the accounting systems.

SUMMARY CONCLUSIONS

Of the six accounting systems discussed in this paper, only the NPV system and the IRR system take account of the cost of holding capital. Just as only these two accounting systems are appropriate for pricing the policy, only these two accounting systems provide measures of profitability that are suitable for monitoring management performance.

The IRR and NPV accounting frameworks provide the same "accept or reject" decision for policy pricing. However, they recognize profits differently.

- A net present value analysis recognizes profits when the project is undertaken.
- An internal rate of return analysis recognizes profits ratably over the lifetime of the project.

The following characteristics apply to prospective analyses, or the expected economic value added in each policy period.

NPV: In the NPV accounting system, the capital at policy inception is the present value of future implied equity flows. In subsequent periods, the net income is just offset by the dollar cost of capital, which is the percentage cost of capital times the amount of required capital. The EVA generated at policy inception is the net present value of the project. The EVA in subsequent periods equals the net income minus the dollar cost of capital, which nets to zero.

IRR: The IRR profit recognition framework, which recognizes profits ratably, provides a constant return on invested capital over the lifetime of the project. The profit in any time period is the length of the time period times this constant return. More rigorously, the return in any period of length "t" is $(1+\text{IRR})^t - 1$.

The IRR and the NPV frameworks have the following characteristics:

- At the time of policy inception, the NPV accounting system recognizes all expected profits. The profit recognized in the IRR accounting system at policy inception is zero, since the length of this period is just the moment of policy writing.
- In subsequent periods, where the NPV accounting system shows zero EVA, the IRR accounting system shows a constant IRR. If the IRR is greater than the cost of capital, the EVA is positive; if the IRR is less than the cost of capital, the EVA is negative.

The accounting exhibits in this paper separate the return at policy inception from the return during the first year. From this perspective, the NPV accounting system shows a constant return equal to the cost of equity capital after policy inception. In practice, the return on policy inception is included with the first year return. The IRR accounting system shows the same rate of return in each year, whereas the NPV accounting system does not.

Neither the NPV perspective nor the IRR perspective is inherently superior. We discuss the relative benefits of the two perspectives further below.

ECONOMIC VALUE ADDED

We use an economic value added (EVA) framework for performance measurement and management compensation purposes. The comments above may give the impression that under the NPV accounting system, the EVA is zero for all periods after policy inception, and that under the IRR accounting system, the EVA is a constant percentage of capital for all periods after policy inception. Were this true, the EVA yardstick would have limited use for performance measurement.

The generalizations above apply to prospective pricing analysis. Performance measurement deals with actual performance. Actual performance may be viewed as the combination of expected performance and the variance of actual from expected in each accounting period.

For performance measurement, we restate the principles as follows:[2]

- Under the NPV accounting system, the EVA at policy inception is the NPV of the project. During subsequent periods, the EVA is zero if actual experience exactly matches expected experience. If actual performance is more favorable than expected, the EVA is positive. If actual performance is less favorable than expected, the EVA is negative.

- Under the IRR accounting system, the EVA at policy inception is zero. During subsequent periods, the EVA is a constant percentage of the required capital at the beginning of the period. It equals the required capital times the difference between the IRR and the cost of equity capital, as long as actual experience matches expected experience. More favorable experience raises the EVA, and less favorable experience reduces the EVA.

2 Measurement of actual performance is discussed more fully in Schirmacher and Feldblum, "Financial Pricing Models for Property-Casualty Insurance: Retrospective Analysis and Performance Measurement."

Illustration: Managers A and B each manage a $100 million book (in premium volume) of commercial fire business. All policies have effective dates of January 1.

At the start of 20XX, Manager A's book of business has a net present value of $5 million, based on best estimates of future loss costs. Manager B's book of business has a net present value of –$5 million, based on best estimates of future loss costs. Based on our assumptions and expectations, we presume that Manager A is adding value to the company whereas Manager B is reducing the value of the company.

Insurance results are inherently uncertain. The value added to the company depends on actual results, not merely on the assumptions and expectations in policy pricing. If Manager A's book of business suffers unexpectedly large losses more than $5 million greater than anticipated, it will have destroyed company value by the end of the year. If Manager B's book of business is unexpectedly free of large losses (more than $5 million less than expected), it will have added value to the company by the end of the year. The EVA yardstick quantifies the actual value added, not the value that the manager expected to add.

The EVA framework is conceptually identical to the return on capital analysis, but it more clearly separates the return into (i) the net after-tax income to the equityholders and (ii) the cost of equity capital. This enables managers to relate the profit earned from a book of business with the cost of the capital needed to support that book of business.

Although the EVA in each period depends on the accounting system, the present value of the entire series of EVA's equals the NPV of the policy. This is true for all accounting systems, as long as the discount rate for the present value calculation is the cost of capital. A mathematical demonstration of this result is shown in Appendix A.

One may conceive of this relationship as follows. The accounting system moves income and value added from period to period. The accounting system does not enhance or diminish the value added over the lifetime of the project.

- Net income is a nominal amount. The total nominal amount of net income is not affected by the accounting system.
- EVA is a present value concept. The total present value of the EVA's is not affected by the accounting system.

GAAP AND EVA

In other industries, GAAP income and GAAP equity are used for EVA analysis. For insurance profitability, GAAP figures are not an adequate measure of profitability.

GAAP reserve valuation: For non-insurance companies, GAAP equity is a proxy for the capital invested in the project and GAAP income is a proxy for the change in the company value during the accounting period. For property-casualty insurance operations, GAAP reserve valuation follows statutory accounting. GAAP equity and GAAP income are not reasonable proxies of net worth or the change in net worth.

> *Illustration:* An insurer writes large dollar deductible workers' compensation business, with an expense ratio of 35% and a nominal loss ratio of 135%. The present value at policy inception of workers' compensation LDD losses is 45% of their ultimate value. No losses are expected to be paid for several years after policy inception.
>
> If no capital were required to support the policy, the net present value of the policy would be
>
> $$1 - (35\% + 45\% \times 135\%) = 4.25\% \text{ of the policy premium.}$$
>
> Suppose the insurer begins the year with $40 million in capital, it writes $100 million of business on January 1, 20XX, and it earns a yield of 8% on invested funds. Its expected net worth at the end of the year (ignoring taxes) is
>
> ($40 million + $4.25 million) × 1.08 = $47.79 million. Its total income is $7.79 million.
>
> The company's combined ratio is 135% + 35% = 170%. The investment income earned during the year is [$40 million (capital) + $65 million (net premium)] × 8% = $8.40 million. The GAAP equity at the end of the year is
>
> $40 million – 70% × $100 million + $8.4 million = ($21.60) million.
>
> Its GAAP income is –$21.6 million – $40 million = –$61.6 million.

The figures used in the illustration are reasonable estimates for LDD workers' compensation business with a high deductible. Large dollar deductible business has an unusually long tail, but the same effects occur in all commercial liability lines of business.

Capital Requirements

For evaluating the economic value added, other industries often use GAAP statements with adjustments to bring the reported amounts closer to fair market value amounts.[3] The use of "adjusted GAAP" statements with reserves shown at fair value does not solve the problems for property-casualty insurance companies. We analyze such fair value accounting systems further below, showing the profit recognition pattern. Our concern here is the intuition.

Illustration: We continue the scenario above. Suppose that the capital requirements for the year of writing the book of business equal 25% of the written premium plus 15% of the loss reserves. The company's statutory surplus declines by $70 million during the year, less the investment income earned during the year. Based on the capital requirements for written premium and nominal reserves, it needs 25% × $100 million + 15% × 135% × $100 million = $45.25 million of surplus at the end of the year. With an 8% per annum investment yield, it must begin the year with surplus of

$$(\$70 \text{ million} + \$45.25 \text{ million} - \$65 \text{ million} \times 8\%) \div 1.08 = \$101.90 \text{ million.}$$

The cost of supplying this capital is large. We assume that the cost of holding capital equals the cost of equity capital minus the after-tax investment yield.[4] If the cost of equity capital is 6 percentage points above the investment yield, the cost of holding capital is 65% × 8% + 6% = 11.20% per annum.

The total profit from the block of business from a fair value accounting perspective is about $4.25 million. In fact, the company loses about 11.2% × $101.9 million = $11.41 million in the first year alone from the mandated capital requirements. Using the fair value of loss reserves does not solve the double taxation problem on invested capital.

These two issues – the need to hold full value loss reserves and the capital requirements imposed by regulatory authorities – are leitmotifs of this paper. Just as these two items compel us to use pricing models which examine the return on invested capital, they compel us to use profit recognition systems that account for the invested capital.

For insurance enterprises, a useable EVA measure starts with after-tax net income based on an NPV or an IRR accounting system. An EVA measure based on GAAP, statutory, tax, or fair value accounting does not align the performance measure with the actual profitability of the book of

---

3 Common examples of such GAAP adjustments for other industries include

- accounting for depreciation, if GAAP does not use a realistic depreciation basis,
- accounting for investments in subsidiaries, if GAAP ascribes too much value to goodwill,
- accounting for post-retirement benefits, if GAAP spreads liabilities over a period of years.

The most significant adjustment is for the market value of fixed assets, which may differ greatly from cost.

4 See Feldblum and Thandi [2002], "The Target Return on Capital," for alternative perspectives on the cost of holding capital.

business. The choice of the NPV vs the IRR accounting system reflects the perspective on profit recognition – at policy inception or evenly over the policy lifetime.

ACCOUNTING SYSTEMS

We show the profit recognition pattern for the illustration in the companion paper, "Modeling the Equity Flows," (and repeated below) under six accounting systems:

22. Statutory accounting
23. GAAP
24. Tax accounting
25. Fair value accounting
26. Net present value analysis
27. Internal rate of return analysis

For each accounting system, we consider after-tax net income, the reported capital, the return on capital, and economic value added. We sum up the profit recognition patterns as follows:

- GAAP and statutory accounting defer the recognition of income.
- Tax accounting, fair value accounting, and NPV accounting recognize income up-front.
- IRR accounting recognizes income ratably over the policy lifetime.
- Only NPV and IRR accounting take into account the cost of holding capital.

Illustration: The indicated premium for a long-tailed workers' compensation book of business is $100 million. Because of a hard market, the company writes the policies for $105 million.

GAAP and tax accounting show losses in the initial policy year and profits afterwards. Tax accounting shows the $5 million profit during the policy year. The fair value and NPV accounting systems show the profit at policy inception. The IRR accounting system shows the additional return ratably over the policy lifetime.

Illustration

We examine the profit recognition pattern for the following illustration under six accounting systems. The implied equity flows for this illustration are shown in Feldblum and Thandi, [2002], "Modeling the Equity Flows."

> *A company writes and collects a $1,000 annual premium on December 31, 20XX. Acquisition expenses of $250 are incurred and paid on that day. Maintenance and general expenses of $150 are incurred and paid evenly over the policy term.*
>
> *The pre-tax investment yield benchmark is an 8% per annum bond equivalent yield (semi-annual compounding). The marginal tax rate on both underwriting income and investment income is 35%.*

Losses are incurred evenly during the policy term and they are paid over several years. To simplify the illustration, we model the losses as if there were two occurrences with ultimate values of $400 each occurring on June 30, 20XX+1, and December 31, 20XX+1. Both losses are paid on December 31, 20XX+3.

The capital requirements are based on the NAIC risk-based capital formula. For this scenario, the capital requirements are 25% of annual written premium plus 15% of loss reserves.

Definitions

The company cash flows and implied equity flows for the illustration are shown in Exhibit ??; the derivations are described in the companion paper. They are based on cash transactions and statutory mandates; they do not depend on the accounting framework used for performance measurement. We show the net income, the reported capital, the return on capital, and the economic value added for this illustration using each accounting framework.

We clarify the meaning of after-tax net income and capital by accounting system.

- *Statutory accounting:* Net income is after-tax statutory income. Direct charges and credits to surplus are included with income. The measure of capital is statutory surplus.
- *GAAP:* Net income is after-tax GAAP income. Direct charges and credits to equity are included with income. The measure of capital is GAAP equity.
- *Tax accounting:* Net income is taxable income minus the tax liability. The measure of capital is statutory surplus adjusted for tax timing differences. The timing differences most relevant to the pricing model are revenue offset, loss reserve discounting, and the deferred tax assets associated with each of these.
- *Fair Value Accounting:* Capital is determined by marking all assets and liabilities to market. If market values are not available, we use present values at a benchmark investment yield. After-tax net income is determined as the change in capital before any capital contributions or shareholder dividends.
- *Internal rate of return:* Capital is defined as the capital invested by equityholders in support of the policy. The net income is the income needed to achieve a constant internal rate of return over the lifetime of the policy. This is analogous to the yield to maturity for a fixed-income security, which is the internal rate of return of the bond's cash flows.
- *Net present value:* The measure of capital is the discounted value of future implied equity flows, using the cost of capital as the discount rate. The net income in all periods except period 0 is the cost of capital times the capital at the beginning of the period. The net income for period 0 is the net present value of the policy at the cost of equity capital.

The implied equity flows and the capital contributed at time t=0 are the same for all accounting systems. The capital contributed at time t=0 is the initial implied equity flow. In the illustration used here, this is the $412.50 contributed by equityholders before policy inception to fund the initial underwriting loss and the initial surplus requirement.

For each accounting system, we show the flow of net income and of capital, using the definitions above. An accounting system is *consistent* if the net income minus the change in capital equals the implied equity flow. As a consequence, the sum of all implied equity flows equals the sum of all after-tax net income flows. All six accounting systems as defined here are consistent. An example of an inconsistent accounting system is one which recognizes only underwriting income but not investment income. Other examples are the following:

- If direct charges and credits to surplus are not included in statutory income, statutory accounting is not a consistent accounting system, because the sum of the equity flows is not equal to the sum of the after-tax net income flows.
- If direct charges and credits to equity are not included in GAAP income, GAAP is not a consistent accounting system.
- If tax exempt income is not included in taxable income, tax accounting is not a consistent accounting system.

We define the after-tax return on capital as the net income during each accounting period divided by the capital at the *beginning* of the period. In the numerical tables, we divide the net income in column "t" by the capital in column "t–1."

For period 0, the net income is the initial underwriting loss for the statutory accounting, GAAP, tax accounting, and fair value accounting. It is the NPV for the NPV accounting system, and it is zero for the IRR accounting system.

Income Recognition by Accounting System

We derive the return on capital and the economic value added for the six accounting systems. We show the pattern of income recognition, and we explain the implications for performance measurement.

STATUTORY ACCOUNTING

Period	0.0	0.5	1.0	1.5	2.0	2.5	3.0
Income	($162.50)	($44.87)	$104.93	$40.57	$40.21	$23.05	$23.48
ΔSurplus	($162.50)	$60.00	($190.00)	$0.00	$0.00	$0.00	($120.00)
Surplus	$250.00	$310.00	$120.00	$120.00	$120.00	$120.00	$0.00
Return	-39.4%	-17.9%	33.8%	33.8%	33.5%	19.2%	19.6%
EVA	($162.50)	($59.45)	$86.85	$33.58	$33.21	$16.05	$16.49

Table 1: Statutory Accounting: Return on Surplus and EVA

The rows refer to (1) statutory net after-tax income, (2) change in policyholders' surplus, (3) policyholders' surplus, (4) return on statutory surplus, and (5) economic value added.

Statutory accounting shows the net income to the company, not the net income to the equityholders. This is true for tax accounting, GAAP, and fair value accounting as well. Net statutory income has the following pattern:

- A large underwriting loss occurs at policy inception, because the equityholders must fund the equity in the gross unearned premium reserves. The initial underwriting loss does not depend on the profitability of the policy.

- For long-tailed lines of business, additional underwriting losses occur during the policy term, because the equityholders must fund part of the full value loss reserves.

- For very short-tailed lines, statutory accounting shows positive income during the policy year if the business has been priced adequately. The magnitude of the profit or loss during the policy term depends on both the rate adequacy and the length of time between premium collection and average loss payment.

- Investment income accrues steadily after policy expiration. For prospective pricing, there is no expected underwriting gain or loss after policy expiration to augment or offset the investment income. Expected net income (i.e., after-tax underwriting + investment income) is positive regardless of the adequacy of policy pricing, since all the underwriting gain or loss occurs during the policy period, when the losses are incurred.

The capital shown on the statutory balance sheet is policyholders' surplus only; it does not include the capital embedded in the loss reserves and unearned premium reserves. The return on statutory surplus is negative at policy inception, because of the initial underwriting loss. During the policy term, the return on surplus is negative for long-tailed lines of business and either low or negative for the shorter-tailed lines of business. The return on surplus turns positive after the policy expires and as losses are settled, because the investment income on the assets supporting the loss reserves is not offset by amortization of any reserve discount.

Illustration – Initial Underwriting Loss

At policy inception, or period 0, the after-tax net income equals

  the expenses incurred (a negative amount)
+ the tax incurred (a negative amount for a tax liability and a positive amount for a tax refund)
+ the change in the deferred tax asset recognized on the statutory balance sheet.

In the illustration, this is

–$250 (expenses) + $17.50 (tax refund) + $70.00 (deferred tax asset) = –$162.50.

The capital is the statutory surplus. Before the policy is written, the statutory surplus equals the $412.50 that must be contributed by equityholders before policy inception to fund the initial underwriting loss and the regulatory capital requirements.

To visualize this, suppose that a company incorporates on December 30, 20XX, and it writes the policy in the illustration on December 31, 20XX. Its owners must supply it with statutory surplus of $412.50 on December 30 so that it can write the policy on the next day.

- The initial underwriting loss is $162.50.
- The capital requirements at policy inception are $250.00.
- The shareholder contribution needed to begin insurance operations is $162.50 + $250.00 = $412.50.
- The after-tax return on surplus in period 0 is –$162.50 ÷ $412.50 = –39.4%.

Illustration – Subsequent Valuation Period

We name the valuation periods by the ending valuation date. For example, period 2.5 extends from time t=2.0 to time t=2.5, or the first half of year 3. The net income in any period equals

  the premiums earned
+ the investment income accrued
– the losses incurred
– the expenses incurred
– the federal income taxes incurred
+ the change in the deferred tax asset

For a prospective pricing model, premiums are earned and losses are incurred during the policy term, not afterwards, assuming the company holds full value loss reserves.

The statutory net income in period 2.5 equals the investment income minus the tax liability plus the change in the deferred tax asset. The deferred tax asset declines from $33.60 at December 31, 20XX+2, to $16.80 at June 30, 20XX+3, for a change of –$16.80.

$35.46 (investment income) –(–$4.39) (tax refund) +(–$16.80) (change in DTA) = $23.05.

The statutory surplus is the risk-based capital requirement. For period 2.5, this is 15% of the full value loss reserves: $900 × 15% = $120. The change in surplus is zero, since the reserves do not change during this period. The return on surplus is $23.05 ÷ $120.00 = 19.2%.

Economic Value Added

The economic value added analysis subtracts the cost of capital from each period's net income. The length of period 0 is infinitesimally small, so the EVA for period 0 is the same as the net income in period 0 for each accounting system.

For most lines of business, the statutory EVA remains negative throughout the policy term, unless the policy is overpriced.[5] The size of the negative statutory income during the policy term is greater for long-tailed lines of business and smaller for short-tailed lines of business.

The expected EVA turns positive after the policy expiration date, regardless of the premium rate adequacy. If the policy is priced adequately, the total EVA is zero. Since the EVA before policy expiration is negative, the EVA after policy expiration must be positive. If the policy is inadequately priced, the additional losses are accrued during the policy term, not afterwards.

To clarify the positive EVA, we examine the sources of gain or loss after policy expiration:

28. gain: investment income on the assets backing discounted loss reserves
29. gain: investment income on the assets backing the capital embedded in reserves
30. gain: investment income on statutory surplus
31. gain: tax refund on amortization of the interest discount in tax basis loss reserves
32. loss: federal income taxes on all sources of investment income
33. loss: the dollar cost of capital, or the percentage cost of capital × the amount of capital
34. loss: the decline in the admitted portion of the deferred tax asset

19. The amortization of the tax basis loss reserves equals the risk-free mid-term rate times the beginning discounted reserves. As an incurred loss, it is an offset to taxable income, and it offsets the expected investment income on the assets backing the discounted reserves.

20. Statutory accounting admits only the portion of the deferred tax asset which is expected to reverse in the coming 12 months. Since the IRS loss reserve discount factors are relatively even until the last few years of the loss payout pattern, the change in the admitted portion of the

[5] The policy term is the period during which the premiums are earned. The policy lifetime is the period from policy inception until the claims are settled. The illustration uses a one year policy term with a three year policy lifetime.

deferred tax asset is small; see Appendix A of Feldblum and Thandi [2002], "Modeling the Equity Flows," for explanation of the tax effects.

21. The remaining gains are the full (pre-tax) investment income on the assets backing the tax basis discounted reserves plus the after-tax investment income on the remaining assets held by the company. These exceed the dollar cost of capital.

The statutory EVA is a biased measure of performance. Its expected value is negative for the first year and moderately positive for subsequent periods. It penalizes managers for writing profitable business and rewards their successors.

GENERALLY ACCEPTED ACCOUNTING PRINCIPLES

GAAP for insurance transactions follows statutory accounting, with two major differences.

- Pre-paid acquisition costs are capitalized and amortized over the term of the policy.
- The entire deferred tax asset from IRS loss reserve discounting may be recognized.[6]

GAAP Results

Period	0	0.5	1.0	1.5	2.0	2.5	3.0
Income	$0.00	($109.32)	$40.48	$23.77	$23.41	$23.05	$23.48
ΔEquity	$0.00	($4.45)	($254.45)	($16.80)	($16.80)	$0.00	($120.00)
Equity	$412.50	$408.05	$153.60	$136.80	$120.00	$120.00	$0.00
Return	0.0%	(26.5)%	9.9%	15.5%	17.1%	19.2%	19.6%
EVA	$0.00	($133.37)	$16.69	$14.82	$15.43	$16.05	$16.49

Table 2: Generally Accepted Accounting Principles: Return on Equity and EVA

The rows refer to (1) net after-tax GAAP income, (2) change in GAAP equity, (3) GAAP equity, (4) return on GAAP equity, and (5) economic value added.

6 Other SAP-GAAP differences are relevant in specific scenarios; these include

- The timing of the recognition of policyholder dividends for workers' compensation carriers; see SSAP No. 65, "Property and Casualty Contracts."
- The statutory offset for agents' balances more than 90 days past due versus the GAAP offset for uncollectible receivables. This is relevant for commercial casualty policies where the insurer collects the premium periodically over the policy lifetime. See SSAP No. 6, paragraph 9.
- The statutory non-admitted charge of 10% of earned but unbilled premiums vs the GAAP offset for uncollectible receivables. This is relevant for policies with low deposit premiums and high audits. See SSAP No. 53.
- The statutory non-admitted charge of 10% of accrued retrospective premiums vs the GAAP offset for uncollectible receivables. This is relevant for retrospectively rated policies. See SSAP No. 66.

Other SAP-GAAP differences that my be relevant for specific companies include (i) the valuation of subsidiaries, (ii) the post-retirement benefit liabilities, (iii) the valuation of fixed-income securities, and (iv) the statutory Schedule F provision for reinsurance.

GAAP shows more rapid recognition of underwriting income during the policy term than statutory financial statements show, giving a more even spread of income over the policy term. GAAP statements do not show an initial underwriting loss. The GAAP matching principle implies that net underwriting income should be constant over the policy term.[7]

For long-tailed lines of business, the GAAP income during the policy term is low or negative, since loss reserves are set at undiscounted values. The GAAP division of income between the policy term and subsequent valuation periods is distorted almost as much as the statutory income. The primary difference between the two accounting systems is that GAAP statements recognize the full deferred tax asset. This speeds up the recognition of income.

The GAAP deferred tax asset or liability from revenue offset depends on the relationship between the deferred policy acquisition costs (DPAC) and the 20% tax assumption.

- If the GAAP DPAC ratio to written premium is more than 20%, the company sets up a deferred tax liability for 35% of the excess over 20%.
- If the GAAP DPAC ratio to written premium is less than 20%, the company sets up a deferred tax asset for 35% of the amount below 20%.

In the illustration, the deferred policy acquisition costs are 25% of written premium. The GAAP deferred tax liability at policy inception is 35% × (25% – 20%) × the written premium, or

$$35\% \times 5\% \times \$1000 = \$17.50.$$

This equals the tax refund at policy inception. If the DPAC is less than 20% of written premium, the GAAP deferred tax asset equals the tax liability at policy inception.

Total GAAP income during the policy term exceeds total statutory income by the amount of the deferred tax asset from IRS loss reserve discounting that is not recognized on the statutory balance sheet. This difference reverses (relatively smoothly) between the policy expiration date and the final settlement of losses.

[7] For simplicity, the $150 of general expenses that are incurred evenly over the policy term are modeled as a cash flow at time t=½; see the companion paper, Feldblum and Thandi [2002], "Modeling the Equity Flows," for the modeling assumptions and the simplifications used in the illustration. If the $150 of general expenses were spread over the policy term, the net income would be –$133.37 + $75.00 = ($58.37) for the first half year and $16.69 – $75.00 = ($58.31) for the second half year. The 6 cent difference stems from differences in the investment income in the two half years. GAAP spreads the expected underwriting income evenly over the policy term. The incidence of the expected investment income is determined by the expected capital requirements and held reserves in each half year.

Illustration – Deferred Tax Asset: The deferred tax asset on the GAAP balance sheet is $39.20 on December 31, 20XX+1, and $33.60 on December 31, 20XX+2. The difference of $5.60 is the amount of the deferred tax asset that reverse in the 12 months between 12/31/20XX+1 and 12/31/20XX+2. This equals the deferred tax asset on the statutory balance sheet on December 31, 20XX+1.[8]

Illustration – Policy Inception: We compare statutory after-tax net income with GAAP after-tax net income at policy inception.

Statutory after-tax net income is a combination of three items: (i) the pre-paid acquisition costs, (ii) the tax liability or tax refund, (iii) the deferred tax asset from revenue offset. GAAP after-tax net income at policy inception has the following characteristics:

- The pre-paid acquisition costs are capitalized at policy inception. No premiums have been earned and no investment income has accrued, so pre-tax net income is zero.
- Since pre-tax net income is zero, the after-tax net income is also zero. Any tax liability or tax refund is offset by a deferred tax asset or liability of the same size but opposite sign.

We compare the statutory and GAAP figures at policy inception to highlight the differences. The current tax liability is a $17.50 refund at policy inception for all accounting systems.

- Statutory accounting shows incurred expenses of $250 and a deferred tax asset of $70.
- GAAP shows incurred expenses of $0 and a deferred tax liability of $17.50.

The table below shows the current tax liability, the deferred tax liability, and the incurred expenses for GAAP and statutory accounting at policy inception. The statutory deferred tax asset is shown as a negative liability.

	Statutory	GAAP	Difference
Current Tax Liability	($17.50)	($17.50)	$0.00
Deferred Tax Liability	($70.00)	$17.50	$87.50
Incurred Expenses	$250.00	$0.00	($250.00)
Total	$162.50	$0.00	($162.50)

The statutory liability is $162.50 greater than the GAAP liability, so the GAAP after-tax net income is $162.50 greater than the statutory income.

Illustration A: Subsequent Valuation Period: We compare GAAP and statutory accounting for period 2.5 (from time t=2.0 to time t=2.5).

- The policy has expired. The GAAP DPAC is zero, and the deferred tax asset from revenue offset is zero for both GAAP and statutory accounting.

[8] The computation of GAAP and statutory deferred tax assets is explained in Appendix A of the companion paper, Feldblum and Thandi, [2002], "Modeling the Equity Flows." For the long-tailed lines of business, the full deferred tax asset stemming from IRS loss reserve discounting is considerably larger than might be inferred from the illustration here.

- A year or less remains until losses are settled. The deferred tax asset from IRS loss reserve discounting is the same for statutory as for GAAP financial statements.
- Loss reserves are held at full values on both the statutory and GAAP balance sheets.

There are no differences in the last year between GAAP and statutory accounting. The net income, return on capital, and economic value added are the same.

Illustration B: Subsequent Valuation Period: For period 1.5 (from time t=1.0 to time t=1.5) we calculate the GAAP after-tax net income in two ways: (i) by comparison with statutory income and (ii) by an independent computation.

Statutory income is $40.57. The only difference between statutory and GAAP after-tax net income is the change in the deferred tax asset stemming from IRS loss reserve discounting.

- At December 31, 20XX+1, the statutory DTA is $5.60 and the GAAP DTA is $39.20.
- At December 31, 20XX+2, both the statutory and GAAP DTA's are $33.60.
- During 20XX+2, the statutory DTA rises by $28.00, and the GAAP DTA declines by $5.60. The difference is $28.00 – (– $5.60) = $33.60.
- This difference is spread evenly between the two halves of the year, for a difference of $16.80 between statutory and GAAP after-tax net income each half year.
- The statutory after-tax net income in period 1.5 is $40.57.
- The GAAP after-tax net income is period 1.5 is $40.57 – $16.80 = $23.77.

Alternatively, we work out GAAP after-tax net income independently. The after-tax net income equals the investment income minus the liability minus the change in the deferred tax asset:

$36.58 *(investment income) – ($12.80) (tax liability on investment income) – (–$2.80) (tax refund on reserve amortization) + (–$2.80) (change in GAAP DTA) = $23.77.*

For long-tailed lines of business, the GAAP EVA is relatively constant and negative during the policy term, since premiums are earned evenly, losses and incurred evenly, and GAAP expenses are written off evenly over the policy term. The negative sign of the EVA stems from the requirement to hold full value loss reserves.

After the policy expiration date, the GAAP EVA is lower than the statutory EVA, since the full DTA from loss reserve discounting has been recognized during the policy term.

Performance Measurement

None of these four measures – statutory income, GAAP income, statutory EVA, and GAAP EVA – reflects the flow of income to the equityholders, and none is suitable for performance measurement. All of them show losses at policy inception (and during the policy term for long-tailed lines of business) and gains in valuation periods after policy expiration. This is the opposite of the profit recognition pattern needed for effective performance management.

Nevertheless, some insurers use GAAP or statutory income measures for performance measurement, such as "net operating income after tax." This measure does not consider the cost of holding capital, and it has perverse effects on management incentives.

- The more capital that is allocated to a manager, the greater is the investment income, and the higher is the net operating income after tax.
- In truth, the more capital that is allocated to a manager, the lower is the net present value or the internal rate of return of the block of business.

When net operating income after tax is used as the performance measure, operating managers seek more capital for their business units. Some pricing actuaries conclude that the manager does not appreciate the cost of holding capital. This is not necessarily correct. Sometimes the pricing actuary does not appreciate the incentive structure of the performance measurement system. If the performance measurement system does not consider the cost of holding capital, the rational manager acts as though this cost were zero.

TAX ACCOUNTING

Net income is defined as taxable income minus the tax liability. Taxable income is statutory income adjusted for tax timing differences. The timing differences stem from IRS loss reserve discounting and revenue offset, as well as from the deferred tax assets associated with each.

Period	0.0	0.5	1.0	1.5	2.0	2.5	3.0
Income	($32.50)	$37.33	($0.87)	$26.57	$10.21	$39.85	($55.72)
ΔSurplus	($32.50)	$142.20	($295.80)	($14.00)	($30.00)	$16.80	($199.20)
Surplus	$380.00	$522.20	$226.40	$212.40	$182.40	$199.20	$0.00
Return	-7.9%	9.8%	-0.2%	11.7%	4.8%	21.8%	-28.0%
EVA	($32.50)	$15.18	($31.32)	$13.38	($2.17)	$29.21	($67.33)

Table 3: Tax Accounting: Return on Capital and EVA

The rows refer to (1) net after-tax taxable income, (2) change in tax-basis equity, (3) tax-basis equity, (4) return on tax-basis equity, and (5) economic value added.

Terminology

Tax accountants often speak of differences between taxable income and GAAP income (or between taxable income and statutory income) as permanent tax differences. This perspective leads to a tax accounting system that is not a consistent accounting system, since the sum of the implied equity flows does not equal the sum of the after-tax net income flows.

The better perspective is that the tax *rate* differs by type of investment, not that taxable income differs by type of investment. Taxable income differs from statutory pre-tax income or GAAP pre-tax income only in the timing of income recognition, not in the amount of income.

Illustration: A property-casualty insurer has $100 million of municipal bond interest income. For insurance companies, the proration provision of the 1986 Tax Reform Act adds 15% of tax-exempt income to regular taxable income.

Common parlance is to say that the insurer has $100 million of statutory income and of GAAP income, but that its taxable income is 15% × $100 million = $15 million in the regular tax environment and 15% × $100 million + 85% × 75% × $100 = $78.75 million in the alternative minimum income tax environment.[9]

Although this is standard terminology, it confuses the tax rate with the amount of income. Tax accountants use the term taxable income to mean income which is subject to federal income taxes. We use the term taxable income to mean the company's income as seen through the lens of tax accounting. The net after-tax income in the regular tax environment is

$100 million – 15% × $100 million × 35% = $94.75 million.

It makes little sense to say the company has $15 million of pre-tax net income and $94.75 million of after-tax net income. Rather, the company has $100 of pre-tax net income with a 5.25% tax rate, or 15% × 35%, applied to this income.[10]

Illustration – Policy Inception: At policy inception, the pre-tax net income equals the tax basis deferred policy acquisition costs – or 20% of written premium – minus the actual pre-paid acquisition costs. The after-tax net income equals the pre-tax net income minus the tax liability or plus the tax refund:

–$250.00 (expenses) + $200 (tax basis DPAC) + $17.50 (tax refund) = –$32.50.

The capital at time t=0, right after the policy has been written, equals

statutory surplus + the tax basis DPAC – the statutory deferred tax asset, or

---

9 For personal taxpayers and for corporate taxpayers other than insurance companies, municipal bond income is tax exempt. For insurance companies, both life insurers and property-casualty insurers, 15% of this tax exempt income is an offset to the loss reserve tax deduction by the proration provision of the 1986 Tax Reform Act. This effectively imposes a tax rate of 15% × 35% = 5.25% on municipal bond income.

In the alternative minimum tax environment, 75% of the tax exempt income is added to the regular taxable income to obtain the alternative minimum taxable income. The 75% of the tax exempt income is called the ACE adjustment. This terminology is unfortunate, since it appears that the amount of taxable income differs between the regular tax environment and the alternative minimum tax environment, when in truth it is the tax rate on municipal bond income that differs between the two environments.

The alternative minimum tax rate is 20%. The alternative minimum tax rate on municipal bond income is 20% × 15% (the regular taxable income portion) + 20% × 85% × 75% (the ACE adjustment portion) = 15.75%.

10 This view of tax accounting accords with the general accounting perspective in this paper. In no accounting system does income just disappear. The total after-tax net income over the life of the project is the same for all accounting systems; it equals the total cash inflow to the company. The IRS can change the tax rates, and it can vary the tax rate by type of income. It does not change the income received by the company. Similarly, it varies the tax rate between the regular tax environment and the alternative minimum tax environment; the income of the company does not depend on the tax environment.

$250.00 (statutory surplus) + $200 (tax basis DPAC) – $70.00 (DTA) = $380.00.

The return on capital at policy inception is the net income divided by the capital provided by equityholders before policy inception, or –$32.50 ÷ $412.50 = –7.9%.

Illustration – Subsequent Periods: After policy expiration,

the after-tax net income = the statutory after-tax net income

41. the tax basis incurred loss
42. the change in the statutory deferred tax asset.

The tax basis incurred loss equals the amortization of the IRS loss reserve discount.[11]

		Period 2.5	Period 3.0
	statutory net income	$23.05	$23.48
–	tax basis incurred loss	$0.00	$96.00
–	change in deferred tax asset	($16.80)	($16.80)
=	taxable net income	$39.85	($55.72)

[11] For simplicity, the illustration shows the tax basis incurred loss at year-end dates and the change in the deferred tax asset spread through the year. For consistency, an actual pricing model would show both at year-end dates or both spread through the year.

NET PRESENT VALUE

Period	0.0	0.5	1.0	1.5	2.0	2.5	3.0
Income	($62.49)	$20.41	$27.71	$12.13	$10.47	$8.74	$7.90
Δ Capital	($62.49)	$125.28	($267.22)	($28.44)	($29.74)	($14.31)	($135.58)
Capital	$350.01	$475.29	$208.07	$179.62	$149.89	$135.58	$0.00
Return	-15.1%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%
EVA	($62.49)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00

Table 4: NPV Accounting: Return on Capital and EVA

The rows refer to (1) net after-tax NPV income, (2) change in NPV capital, (3) NPV capital, (4) return on NPV capital, and (5) economic value added.

The NPV to the equityholders of a financial project is the present value of the implied equity flows. The discount rate is the cost of equity capital. The NPV is an estimate of the market value of the project.

If the company cash flows are a reasonable proxy for the implied equity flows, we can use the present value of these cash flows. For property-casualty insurers, we must independently estimate the implied equity flows.

The NPV of a project is similar to the fair value of a project, with two differences.

- The fair value accounting system takes the present value of the company cash flows, with property-casualty reserves discounted at the company's benchmark investment yield. The NPV accounting system uses the present value of the implied equity flows, discounted at the cost of equity capital.
- The fair value accounting system does not include the cost of holding capital. It is the accounting system implicit in the consumer's value perspective described in the companion paper ("Modeling the Equity Flows"). The NPV accounting system is the shareholders' cost perspective; the cost of capital is paramount.

The net present value accounting framework recognizes the expected gains and losses at policy inception. This is the perspective used for valuing financial securities. In contrast, an IRR perspective recognizes the expected gains and losses over the life of the contract.

Illustration – Common Stock: Financial valuation models assume that the market value of a share of common stock equals the present value of the expected future dividends. A change in the expected future dividends causes an immediate change in the market value. There is no amortization over the expected holding period of the stock. Amortization is an accounting concept, not a financial concept.

Illustration – Fixed-income Securities: The market value of a fixed-income security is the present value of the expected coupons and principal repayment. If market interest rates change after issuance of the fixed-income security, market value accounting recognizes the change in value immediately.

NPV Income and EVA

Four principles govern the NPV net income and economic value added:

- The expected income at policy inception equals the NPV of the project.
- The expected income in valuation periods after policy inception is the dollar cost of capital, or the percentage cost of capital × the capital held at the beginning of the period.
- The expected economic value added at policy inception equals the NPV of the project.
- The expected economic value added in valuation periods after policy inception is zero.

The NPV framework implicitly assumes that the business risk in managing a policy after it has been written is not material compared to the underwriting and pricing risk at policy inception. This is a fair assumption for most personal lines products, such as life insurance, annuities, automobile insurance, and Homeowners. When there is continuing company involvement in the management of risk, such as loss engineering services or managed care rehabilitation services, the NPV analysis perhaps allocates too much of the gain or loss to the policy inception date.

A performance measurement system uses actual income, not expected income. The actual income in valuation periods after policy inception equals (i) the expected income (ii) plus or minus unexpected gains or losses. These unexpected gains or losses may result from

- random loss fluctuations
- unanticipated reserve development
- changes in financial conditions, such as interest rates or common stock prices.[12]

PERFORMANCE MEASUREMENT

The IRR and NPV analysis are alternative perspectives which give the same "accept or reject" decision for the proposed project.

- The NPV is better suited for rate filings, since a loss of value is denoted by a negative dollar figure. If the project is not profitable, the IRR is less than the cost of capital, but it may still be positive; see Feldblum [1992: IRR].
- The IRR is better suited for comparison of projects, since it is not affected by the size or the duration of the project.

For performance measurement and the recognition of income, the two methods differ.

---

[12] This difference between expected and actual income is true for all accounting systems. We use the NPV the IRR accounting systems for retrospective performance measurement; see Schirmacher and Feldblum [2002] "Retrospective Analysis and Performance Measurement." GAAP uses similar "true-up" procedures for determining the profitability of universal life-type contracts; see SFAS 97.

- The IRR determines a constant rate of return for the equityholders. Income is recognized ratably over the lifetime of the project. An IRR of 10% per annum implies that the equityholders earn 10% each year on their invested capital.
- The NPV provides a single dollar figure, generally valued at inception of the project. An NPV of $5,000 implies that the business decision to undertake this project has added $5,000 to the company's worth. If the return at policy inception is added to the first year return (as is done in practice), the return on capital differs between the first year (the policy term) and all subsequent years.

Illustration: A workers' compensation policy with a 20 year lifetime until final settlement of losses is sold on January 1, 20XX for $500,000. At a 12% per annum cost of equity capital, the policy has an NPV of $50,000. The internal rate of return is 14% per annum.

The NPV analysis recognizes the $50,000 profit on January 1, 20XX. This accords with economic reasoning, since the profit results from the underwriting decision made on that day. The nominal after-tax net profit in subsequent periods is 12% times the invested capital, or a $0 economic value added. This perspective is not consistent with the conceptual accounting principles underlying GAAP, which recognize profit over the course of the contract.[13]

The IRR analysis recognizes 14% of the initial capital contribution as the net income in 20XX, 14% of the capital requirement in 20XX+1 as the net income in 20XX+1, and so forth. Since the cost of equity capital is 12% per annum, this represents a 2% economic profit each period. This economic profit is termed here the EVA. This perspective accords with the GAAP principle of recognizing income ratably until performance is complete, but it does not accord with economic reality. There may have been no business action in 20XX+1 and subsequent years that warrants the recognition of profits in those years.

COMPARISON WITH OTHER ACCOUNTING SYSTEMS

A comparison of the NPV accounting system with the three systems discussed earlier (statutory, GAAP, and tax accounting) clarifies the important attributes.

Statutory accounting and property-casualty insurance GAAP defer the recognition of profits, since both accounting systems use undiscounted loss reserves and statutory accounting does not recognize deferred policy acquisition costs as an asset.[14]

---

13 Actual GAAP for property-casualty insurance contracts values loss reserves on an undiscounted basis, just as statutory accounting does, so it defers the recognition of profits over the settlement period of the losses. GAAP for life insurance contracts illustrates the recognition of profits over the course of the policy. For traditional policies, profits are recognized ratably over the premium paying period; see SFAS 60. For universal life-type contracts, net profits are recognized in proportion to expected gross profits; see SFAS 87. For retroactive reinsurance contracts, GAAP recognizes the profit over the settlement period of the reserves; see SFAS 113. GAAP for life insurance has a slight conservative tilt (conservative interest rates and conservative mortality assumptions) that defer some of the profit to later periods.

14 Differences in expense costs and expected loss costs between new and renewal policies magnifies the deferral of profits for property-casualty insurance contracts under all accounting systems. This paper prices a single policy, using the average loss costs and expense ratios. A more complete analysis is needed for pricing a cohort of policies; see Feldblum [Asset Share, 1996].

Tax accounting uses discounted loss reserves and it recognizes deferred policy acquisition costs by the revenue offset provision. It shows the same profit recognition pattern as the NPV analysis. The dollar amounts may be slightly different, since the revenue offset provision may not exactly match the deferred policy acquisition costs and the IRS loss reserve discounting calculation may not exactly match an economic discounting calculation.[15]

NPV vs EVA

The net present value is a "stock" of money that represents the worth of a project; it is generally valued at inception of the project. The EVA is a "flow" of money that represents the value added in each reporting period. The NPV is the present value of the future EVA's, discounted at the cost of capital.

The NPV calculation combines the net after-tax income with the required capital flows and determines the worth of a project. The EVA calculation separates the net after-tax income from the cost of holding capital. This facilitates the reconciliation of the EVA calculations with the company's accounting statements, and it allows managers to more easily understand the performance measurement system.

The EVA in an accounting system reflects the attributes of that accounting system. To illustrate this, we contrast the EVA based on statutory accounting with the EVA based on NPV accounting.

Because of the requirements to hold gross unearned premium reserves and undiscounted loss reserves, statutory accounting shows negative expected income the first year of a policy's life and positive income in subsequent years (assuming the book of business is profitable). EVA magnifies this distortion because the required capital is generally greatest during the first year of the policy's lifetime, after which the required capital declines.

For the GAAP, statutory, tax, and fair value accounting systems, the EVA equals the after-tax net income minus the dollar cost of holding capital. These accounting systems have accepted definitions of net income. The NPV accounting system is a financial accounting system. It is the accounting system that supports the recognition of profit as the net present value of the project. This means that the EVA equals the NPV at policy inception and is zero in all other periods.

We can rephrase this by defining the performance measurement yardstick as the change in the discounted value of future EVA from the beginning of the year to the end of the year and subtracting the cost of holding capital. This measure recognizes all profit up front – at the time of the underwriting decision or the sale of the insurance contract. Inadequate premiums cause a negative expected profit at policy inception; redundant premiums cause a positive expected profit at policy inception. The expected profit in this measure is zero for all subsequent years, regardless of the premium rate level.

ILLUSTRATIONS

---

15 In the simplified illustration used here, all $150 of general expenses are recognized at time t=½; in practice, these expenses would be recognized evenly over the policy term. In addition, the illustration here uses a 2 to 2½ year lag between loss occurrence and loss payment, but it uses IRS loss reserve discount factors that are appropriate for a line of business with a longer lag. These simplifying assumptions ease the calculations, but they slightly distort the expected income recognition patterns. The comments in the text refer to the patterns observed in actual blocks of business, not in the illustration here.

The NPV is the present value of the implied equity flows at policy inception, discounted at the cost of equity capital. In truth, any date may be chosen as the valuation date for the NPV perspective; we use the policy inception date by convention. This implies that we recognize all expected profits and losses up front, at the time of the business decisions that lead to the expected profit or loss.

The capital at any point in time is the economic worth of the project, or the present value of future implied equity flows. The illustrations below use an annual effective rate of 12% as the cost of capital, or 5.83% for each half-year valuation period ($1.0583^2 = 1.120$). The net income in any period equals the implied equity flow plus the change in required capital (cf Robbin 1993).

Illustration: We begin with period 3.0, the final valuation period. The cost of equity capital is 12% per annum, or 5.83% each half year. The implied equity flow at time t=3.0 is \$143.48. The capital at the beginning of period 3.0 is \$143.48 / 1.0583 = \$135.58. The net income in period 3.0 is \$143.48 – \$135.58 = \$135.58 × 5.83% = \$7.90.

The capital at the end of period 2.5 is the capital at the beginning of period 3.0 plus the implied equity flow at time t=2.5:

\$135.58 + \$23.05 = \$158.63.

The capital at the beginning of period 2.5 is \$158.63 discounted for half a year at the cost of equity capital, or \$158.63 / 1.0583 = \$149.89. The net income for period 2.5 equals the implied equity flow plus the change in capital, or

\$23.05 + (\$135.58 – \$149.89) = \$8.74.

We continue in this fashion back to the earliest valuation date.

The EVA for all periods after policy inception is zero, since the net income equals the capital at the beginning of the period times the cost of capital. The EVA at policy inception is the NPV of the project.

When the company's business risk occurs predominantly at policy inception, the NPV performance measurement system is ideal. Business risk is the risk stemming from the company's business decisions. For most policies, the predominant business risk lies in the underwriting decision to accept or reject the policy and in the choice of the premium rate.

Some insurance risks do not occur at policy inception. For instance, risks relating to random loss fluctuations or unanticipated loss development do not occur at policy inception.

Illustration: A catastrophe excess-of-loss treaty on January 1, 20XX covers the layer \$40 million excess of \$10 million, with a rate of 3% of subject premium. The underwriting and pricing risks on January 1, 20XX, are the reinsurer's decision to write the treaty and the chosen premium rate. For a first dollar book percentage point business, *these underwriting and pricing risks account for the majority of the variance in the final profit or* loss. The NPV profit recognition perspective is suitable. For the catastrophe reinsurance treaty, the random occurrence of windstorms and other natural catastrophes is a material risk. For the reinsurance treaty, an IRR profit recognition pattern better reflects the "release from risk."

Some business risks, such as the legal defense stance on complex tort liability cases, are not always within the purview of underwriting managers responsible for line of business profitability. Performance measurement for in-house legal counsel would not be measured by an NPV pattern that recognized profits on the policy inception date.

We note these exceptions to guard against over-generalization of the profit recognition pattern exemplified by the NPV accounting system. For most insurance operations, the NPV perspective is appropriate.

FAIR VALUE ACCOUNTING

Period	0	0.5	1	1.5	2	2.5	3
Income	($37.88)	$14.98	$19.78	$8.77	$7.50	$6.19	$5.52
Δ Capital	($37.88)	$119.85	($275.15)	($31.80)	($32.71)	($16.85)	($137.96)
Capital	$374.62	$494.48	$219.33	$187.52	$154.82	$137.96	$0.00
Return	-9.2%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%
EVA	($37.88)	($6.86)	($9.05)	($4.01)	($3.43)	($2.83)	($2.52)

Table 5: Fair Value Accounting: Return on Capital and EVA

The rows refer to (1) net after-tax FV (fair value) income, (2) change in FV capital, (3) FV capital, (4) return on FV capital, and (5) economic value added.

Fair value accounting values all assets and liabilities at market value if a liquid market exists, or at a discounted value if no liquid market exists. Market values or reasonable proxies exist for most financial assets held by insurers.

For property-casualty loss reserves, liquid markets do not exist; a proxy is needed. Financial theory assumes that the market value for stocks and bonds equals the present value of the expected future dividend payments, coupon payments, and principal repayments. By analogy, the fair value of loss reserves should equal the present value of the expected loss payments.

Actuaries differ on the proper discount rate for loss reserves. The exhibits here use the pre-tax benchmark investment yield, so that assets and liabilities are valued at the same rate. This is similar to the GAAP investment yield benchmark for valuing pension liabilities; see SFAS 87. Non-investable assets and non-traded assets may be valued in the same manner.

We mention below other views on the appropriate discount rate. For more complete discussion, see Feldblum and Thandi [2002], "Benchmark Investment Yield."

56. Woll [1987], followed by Lowe [19??] uses the risk-free rate to give the value of the underwriting operations in an "economic value accounting" framework. The differential between the risk-free rate and the company's investment yield is the value added by the company's investment department, not by its underwriting operations.

57. Myers and Cohn [1987], following on the work of Kahane [1978], Hill [1979], and Fairley [1979], use a CAPM-type adjustment to the risk-free rate to derive the loss discount rate. The beta of losses in the Myers/Cohn model is intended to reflect the covariance of the loss cash flows with the overall stock market returns.

The theory of underwriting betas was popular in the late 1970's and early 1980's, when the CAPM was the dominant stock valuation model for investment analysts. The CAPM has since lost its luster in the financial world, particularly after the 1992 Fama and French analysis of stock market anomalies. In addition, the betas of losses has been impossible to measure or even detect (see Cummins and Harrington [1984], Kozik [1995], and Feldblum [1996: Betas]). The theory of underwriting betas is rarely mentioned now except in Massachusetts rate hearings. See Feldblum, [PCAS d/d disc].

58. Butsic [1988] uses a utility theory argument to determine the economic value of reserves, and he suggests a 3% to 4% risk adjustment below the risk-free rate. Although popular in the casualty actuarial community, Butsic's perspective does not differentiate between systematic risk and unique risk.

We do not arbitrate among these views here. Our objective is to show the accounting flows and the pattern of income recognition.

The selected discount rate affects the pattern of profit recognition. A higher discount rate causes profits to be recognized earlier; a lower rate causes profits to be recognized later (cf Lowe and Philbrick [1985], Lowe [19??: GAAP]).

Illustration – Policy Inception: The cash outflow at policy inception equals the pre-paid acquisition costs plus the tax liability: $250 + (–$17.50) = $232.50. The present value at a 8% per annum bond equivalent yield of the future losses, expenses, and taxes is $805.38. The premium is $1,000, and the net income equals

$1,000.00 – $232.50 – $805.38 = –$37.88.

Some version of fair value accounting may eventually be adopted for international accounting standards and perhaps even by U.S. regulatory authorities. We summarize several of the relationships with other accounting systems.

Fair value accounting is like an NPV accounting system in that the capital at any time is the present value of future losses and expenses. The specifics of the computation are different.

- The fair value accounting in this paper uses a benchmark investment yield to determine the present value of insurance liabilities. The NPV accounting system uses the cost of equity capital as the discount rate.

- Fair value accounting takes the present value of the insurance cash flows. The NPV accounting system takes the present value of the implied equity flows. The fair value perspective is similar to the consumer's value perspective discussed in Feldblum and Thandi, [2002], "Modeling the Equity Flows," since it takes no account of the cost of holding capital.

Some actuaries argue that the fair value assessment is only part of profitability measurement, and that explicit account must be taken of the cost of holding capital. We agree with this view; this makes the fair value perspective similar to the NPV perspective, though the hurdles rates are different. Other actuaries add that a negative risk adjustment to the loss reserves discount rate, as proposed by adherents of underwriting betas, transforms the fair value perspective into the net present value perspective. Numerically, there is some truth in this.[16] We prefer to discuss the two accounting systems separately, since the rationale for each is different.

FAIR VALUE ACCOUNTING AND TAX ACCOUNTING

Fair value accounting is similar to tax accounting. Both value loss reserves at discounted values, and neither takes account of the cost of capital. However, tax accounting uses pre-set formulas, whereas fair value accounting uses actuarial or financial estimates. The following examples compare the treatment of pre-paid acquisition costs and of loss reserve discounting among the accounting systems.

Deferred policy acquisition costs: The revenue offset provision in the Internal Revenue Code assumes that pre-paid acquisition costs equal 20% of written premium. GAAP uses the actual acquisition cost percentage. The revenue offset provision does not necessarily equal the actual pre-paid acquisition costs. Both GAAP and tax accounting spread the profit or loss in the rest of the policy premium over the policy term.

> *Illustration:* If the pre-paid acquisition costs are 20% of written premium, tax accounting has no gain or loss at policy inception. If the pre-paid acquisition costs are 15% of written premium, tax accounting recognizes 5% of written premium as an immediate gain at policy inception. If the pre-paid acquisition costs are 25% of written premium, tax accounting recognizes 5% of written premium as an immediate loss at policy inception.

Instead of a using a deferred policy acquisition cost asset, fair value accounting values the unearned premium reserve as the present value of future losses and expenses. The NPV and IRR accounting systems implicitly do the same. Fair value accounting and NPV accounting recognize the expected profit or loss at policy inception. The IRR accounting system spreads the profit or loss over the lifetime of the policy.

The expected profit or loss from the policy is a subjective estimate. GAAP and tax accounting seek objective estimates. The pre-paid acquisition costs incurred when issuing the policy can be quantified, and GAAP relies on this figure. The categorization of expenses as pre-paid acquisition costs is somewhat subjective, and it relies on the discretion of the insurer. Tax accounting prefers to rely on a strict formula, which cannot be changed by company management. The NPV, IRR, and fair value accounting systems are geared to policy pricing and internal performance measurement, not to external reporting. They rely on internal (actuarial) estimates of policy profitability, not on actual costs incurred.

---

16 Butsic's [1988] formula, $Z = e \times (R - r_f)$, formalizes this relationship between the loss discount rate and the cost of equity capital.

Loss reserve discounting: There is no consensus on loss reserve discounting rates or procedures. GAAP, statutory, and tax accounting seek objective figures. GAAP and statutory accounting use undiscounted reserves unless the discount is derived from a published mortality or morbidity table. The taxing authorities need discounted reserves to speed up the incurral of tax liabilities, but they seek to avoid company discretion in the choice of discounting parameters. Tax accounting uses a 60 month average of federal mid-term rates along with a formulaic loss payment pattern to determine discounted reserves.

Fair value accounting uses the discount rate and the loss payment pattern appropriate for the given book of business, as selected by the pricing actuary. The IRS loss reserve discount factors do not necessarily reflect the true discount in full value reserves.

The NPV and IRR accounting systems also use a company determined loss payment pattern, though neither system discounts reserves. The NPV accounting system determines implied equity flows based on the loss payment pattern, and it discounts the implied equity flows at the cost of equity capital. The IRR accounting system does not discount any figures. It determines the nominal implied equity flows and determines the internal rate of return among them.

Internal Rate of Return

Period	0.0	0.5	1.0	1.5	2.0	2.5	3.0
Income	$0.00	$6.13	$7.77	$3.51	$2.96	$2.41	$2.10
Δ Capital	$0.00	$111.00	($287.15)	($37.07)	($37.25)	($20.64)	($141.38)
Capital	$412.50	$523.50	$236.34	$199.28	$162.02	$141.38	$0.00
Return	0.0%	1.5%	1.5%	1.5%	1.5%	1.5%	1.5%
EVA	$0.00	($17.92)	($22.75)	($10.27)	($8.66)	($7.04)	($6.14)

Table 6: IRR Accounting: Return on Capital and EVA

The rows refer to (1) net after-tax IRR income, (2) change in IRR capital, (3) IRR capital, (4) return on IRR capital, and (5) economic value added.

The IRR is the rate of return which equates the present value of the equity inflows with the present value of the equity outflows. One may also conceive of the IRR as the dividend yield that the equityholders receive each year.[17]

Measuring Base

The IRR perspective shows a constant rate of return as a percentage of the invested capital. Some actuaries presume that the amount of invested capital in an insurance project should reflect the risk of the project. A greater amount of risk requires a greater amount of capital and a greater dollar return.

This perspective is not necessarily correct. The capital invested in an insurance project is based upon statutory mandate, not actuarial risk quantification. This capital comprises two pieces: (i) the capital embedded in statutory reserves, and (ii) the capital explicitly held in policyholders' surplus.

- The capital embedded in reserves is not risk related, except in so far as longer duration reserves are more risky. Some long duration reserves, such as reserves for environmental and toxic tort liabilities, are highly uncertain. Other long duration reserves, such as reserves for long term disability cases and workers' compensation lifetime pension cases, are not more risky than shorter-tailed casualty reserves.

---

[17] The implied dividend each period is appealing to some business managers. The implied dividend is sometimes more and sometimes less than the implied equity flow.

Illustration: The equityholders fund an insurance project with $10,000 at time t=0, and they receive $12,100 at time t=2. The formal definition of the internal rate of return sets $\$10,000 - \$12,100 / (1+z)^2 = 0$ and solves for z = 10%. To conceive of the IRR as a periodic dividend, we assume a dividend payment of $1,000 at time t=1 along with a capital contribution from the equityholders to the company of $1,000 at time t=1. The combined cash flows form two loans at 10% per annum interest: a two year loan of $10,000 at time t=1 and a one year loan of $1,000 at time t=1.

- The capital explicitly held in surplus reflects the risk to *policyholders* that they may not be reimbursed for their losses. This is not the same as the risk to equityholders.[18]

Illustration – Policy Inception: The income recognition pattern for the IRR perspective is determined by the constant return on invested capital. The initial capital contribution is $412.50. At time 0, the return to the equityholders is IRR × 0 years × $412.50 = $0, since the invested capital has been held for an insignificant amount of time.

Illustration – Valuation Periods after Policy Inception: At time t=½, the $412.50 has been held for half a year. The semi-annual internal rate of return computed from the implied equity flows is 1.485%. We derive net income and capital amounts as follows:

- The net income is 1.485% × $412.50 = $6.13.
- The implied equity flow for time t=½ is –$104.87; see the table of implied equity flows at the beginning of this paper.
- The implied equity flow equals the net income minus the change in capital, so the change in capital is $6.13 – (–$104.87) = $111.00.
- The total capital at time t=½ is $412.50 + $111.00 = $523.50.

We use the same procedure for other valuation periods.

- At time t=1.0, the $523.50 has been held for half a year.
- The net income is 1.485% × $523.50 = $7.77.
- The implied equity flow for time t=1.0 is +$294.93; see the table of implied equity flows at the beginning of this paper.
- The change in capital is $7.77 – $294.93 = –$287.15.
- The total capital at time t=1.0 is $523.50 – $287.15 = $236.64.

Alternatively, we could determine the net income and capital amounts starting from the last valuation period.

- The final implied equity flow at time t=3.0 is $143.48.
- Since the return is 1.485% each half year, the capital at the beginning of the final valuation period is $143.48 ÷ 1.01485 = $141.38.
- The net income in the final valuation period is $143.48 – $141.38 = $2.10.

IRR and Yield to Maturity

Insurance risk varies over the lifetime of a block of business. Presumably, the rate of return should be higher when the risk to the equityholders is higher. The risk to equityholders is greatest during the policy term, when there is uncertainty regarding the occurrence of claims.

---

18 The invested capital reflects risk in that the invested capital is the maximum amount that can be lost by equityholders. The variability of gain or loss differs by line of business, but the amount of capital "at risk" is represented by the invested capital. For a more complete discussion of the risk inherent in different lines of business, see Feldblum and Thandi [2003: Capital Allocation].

The IRR perspective makes no such adjustments. The rate of return is a level amount over all periods, just like the yield to maturity is a level yield over the life of the security.

Illustration: Suppose the Treasury spot rates are 5% for one year, 6% for two years, and 7% for three years. For simplicity, assume that these rates are effective annual yields, and that Treasury securities have annual coupon payments. The coupon rate and the yield to maturity "Z" for an on-the-run three year Treasury note issued at par would be the solution to

$$Z/(1.05) + Z/(1.06)^2 + (1+Z)/(1.07)^3 = 1$$
$$Z = 6.91\%.$$

If the Treasury note is issued with a 7% annual coupon, the market value of the note is

$$\$7/(1.05) + \$7/(1.06)^2 + \$107/(1.07)^3 = \$100.24.$$

The yield to maturity of this security is the solution to

$$\$7/(1+Z) + \$7/(1+Z)^2 + \$107/(1+Z)^3 = \$100.24$$
$$Z = 6.909\%.$$

The internal rate of return has the same interpretation. It is the constant yield over the lifetime of the policy that provides the appropriate return to investors for the risk undertaken in each period.

Illustration: Suppose that the capital required to support an insurance policy is $10,000 during the policy term. After policy expiration, the required capital runs off as losses are paid. The amount of capital needed each year is $5,000 in the first year after policy expiration, $3,000 in the next year, and $1,000 in the next year. Investors require a 15% return on capital during the policy term and an 8% return on capital after policy expiration.

The required after-tax net income for each year of this policy and the implied equity flows are shown in the table below.

Year	*Implied Equity Flow*	*Assets*	*Net Income*	*Rate of Return*
1	($10,000)	$10,000	$1,500	15%
2	$6,500	$5,000	$400	8%
3	$2,400	$3,000	$240	8%
4	$2,240	$1,000	$80	8%
5	$1,080			

The internal rate of return "Z" on the implied equity flows is 11.984%. The 11.984% internal rate of return may be thought of as a 15% return on capital during the policy term and an 8% return on capital as the reserves run off. The constant internal rate of return is an accounting construct to simplify the presentation of the policy profitability.

Funding the Rate of Return

The IRR perspective lends itself to an intuitive understanding of the return to equityholders. The income of an insurance enterprise may be divided into two pieces:

- the investment income on equityholder provided capital, and
- the profits from insurance operations.

The sum of these two pieces is the return on invested capital. The profits from insurance operations includes both the underwriting income and the investment income from the policy transaction funds (sometimes termed the policyholder supplied funds).

We use the figures in our illustration. The benchmark investment yield is 8% per annum compounded semiannually, or 4% each half-year. The target IRR equals the cost of equity capital, or a 12% effective annual rate.

The insurance operations must fund the cost of holding capital. The cost of holding capital is the cost of equity capital minus the investment yield received on the equityholder supplied funds. The cost of equity capital is an after-tax return. The benchmark investment yield of 8% is a pre-tax yield. The corresponding after-tax yield is $8\% \times 65\% = 5.2\%$ per annum.

To induce equityholders to provide supporting capital, the insurance enterprise must provide two parts of the required return to equityholders:

- The federal income taxes on the investment income on equityholder funds. This is the cost of double taxation, or the difference between the pre-tax yield and the after-tax yield, or 8% – 5.2% = 2.8% per annum. The equityholders could obtain the pre-tax yield by investing directly on their own (or by investing in a mutual fund), instead of investing through the insurance enterprise.
- The difference between the cost of equity capital and the benchmark investment yield, or about 4% per annum.[19] This is the compensation for the risk undertaken by the equityholders. If the insurance enterprise provided a return just equal to the benchmark investment yield, the equityholders would prefer to invest their money in a mutual fund or directly in the financial markets, thereby avoiding the risk of insurance operations.

These two components of the policy premium comprise the profit margin. The remainder of the premium funds the expected loss and expense costs of the policy.[20]

---

[19] The benchmark investment yield in the illustration is bond equivalent yield, whereas the cost of equity capital is an effective annual yield. On an effective annual basis, the investment yield is 8.16%.

[20] These two components are after-tax amounts. The policy premium is a pre-tax amount. Using the figures in the text, the needed profit margin in the policy premium would be (2.8% + 4%) / (1 – 35%) = 10.46% for each year that equityholder provided capital is invested.

EQUITYHOLDERS AND POLICYHOLDERS: FUNDING THE INSURANCE POLICY

The combination of the IRR perspective and the fair value perspective allows a finer analysis of the funding of the insurance policy.

Were there no capital requirements and no statutory mandates for full value reserves, the "consumers' value perspective," as reflected in a fair value accounting system, would properly price an insurance product.[21] The requirement to hold capital imposes an additional cost, the cost of holding capital.

At a minimum, this cost is the cost of double taxation on the investment income on equityholder supplied funds (see Myers and Cohn [1987]). The pricing model here implicitly assumes that the cost of holding capital is the difference between the cost of equity capital and the after-tax investment yield of the company (cf. Atkinson and Dallas [2000, ch 11]). To induce equityholders to provide funds to support insurance operations, this cost of holding capital must be paid by the policyholders. The tax on underwriting income adds an additional cost to capital funded through the policy premium.

THREE SCENARIOS

---

21 On the consumers' value perspective, see Feldblum and Thandi, [2002], "Modeling the Equity Flows."

To clarify the funding of the insurance policy, we trace the flow of funds for three scenarios, which differ only in the premium rate. For each scenario, the cost of the policy is the present value of all benefits, expenses, and federal income taxes on the insurance transactions. The discount rate for the present value calculation is the pre-tax investment yield.[22]

21. When the internal rate of return on the implied equity flows is less than the pre-tax investment yield, the policyholder premium is not sufficient to pay the costs of the policy, and the equityholders must supply capital to pay the unfunded costs.
22. When the internal rate of return on the implied equity flows exactly equals the pre-tax investment yield, the policyholder premium is just sufficient to pay the costs of the policy. The indicated premium in this scenario is the premium determined in a fair value accounting system that takes no account of invested capital. The equityholders receive the company's benchmark investment yield on their funds, not the cost of equity capital. The return on the invested capital is not sufficient to induce them to supply funds to the insurance industry.[23]
23. When the internal rate of return on the implied equity flows exceeds the pre-tax investment yield, the profit in the policyholder premium is transferred to equityholders to fund all or part of the cost of holding capital. The equityholders will supply capital only if the internal rate of return on the implied equity flows is at least equal to the cost of equity capital.

The policy premium is $1,000.00 in Scenario A, $1,058.27 in Scenario B, and $1,096.07 in Scenario C. A higher policy premium needs a lower capital contribution from equityholders. A lower capital contribution means a lower cost of holding capital. The schematic for each scenario shows the implied equity flows. These schematics address the issue of "who is funding whom" in each scenario.

SCENARIO A: PREMIUM INADEQUACY

If the insurance policy is inadequately priced, the profit margin in the premium is not sufficient to provide the needed return on equityholder funds. The illustration in the text, with a policy premium of $1,000, shows an IRR of 3% per annum, or $(1.03^{0.5} - 1) \approx 1.485\%$ each half year.

This does *not* mean that the insurance operations are earning a 3% profit. The IRR is the net income; it does not subtract the cost of equity capital. The company's economic profit is reflected by the EVA, not by the IRR.

To see this, we compare the insurance enterprise to an investment trust. The equityholder provided capital would earn an 8.0% in a pure investment trust. An investment trust, such as a mutual fund, passes the investment earnings to the investors without having to pay corporate income taxes on the earnings. Unlike the investment trust, the property-casualty insurance enterprise is fully taxable; the

---

22 Since we are explicitly modeling federal income taxes as an expense, we use the pre-tax investment yield as the discount rate, not the after-tax investment yield. For further discussion of pre-tax and after-tax investment yields for modeling insurance operations, see Feldblum, [2002], "The Pricing of Commutations."

23 We include the federal income taxes on the investment income on equityholder supplied funds as a policy cost, since this is a cash outflow from the company stemming from state mandated capital requirements.

after-tax return is 5.2%. Not only is the policyholder premium too meager to reimburse the equityholders for the costs of double taxation, but the premium is not sufficient to cover the losses and expenses of the insurance operations. Part of the 5.2% after-tax return is transferred to policyholders to finance the insurance operations.

If the policy premium were even lower, the investment income on the equityholder funded capital might not suffice to fund the insurance operations. In such a scenario, the IRR on the implied equity flows would be negative, and the equityholders' capital would be invaded to fund the insurance operations.

Exhibit ?? shows a schematic for the first two periods of Scenario A. An implied dividend of 1.485% is paid each half year to the equityholders from the investment income on their funds. The rest of the investment income is transferred to the insurance operations: $10.37 in the first half year and $13.17 in the second half year.

An alternative means of viewing the flow of funds is to conceive of the equityholder provided capital in two parts. At policy inception, the equityholders provide $412.50. Of this amount, $374.62 is used to provide the 1.485% dividend each half year and to repay the principle as the losses are settled. We don't need the full $412.50, because the 1.485% equityholder dividend each half year is less than the after-tax investment yield of 3.85% each half year.

The remainder of the funds, or $37.88, is used to offset the deficiency in the policyholder premium. If the policy were adequately priced, part of the policy premium funds the cost of holding capital. We call this the policyholder funded capital. In this scenario, the policy is inadequately priced, and the schematic labels this $37.88 as the negative of the policyholder funded capital, or "–PFC."

Illustration: The $412.50 of equityholder funded capital accumulates at the pre-tax investment yield, adding $412.50 × 4.0% = $16.49 in the first half year. The equityholder dividend is 1.485% (semi-annual internal rate of return) of the contributed capital, or $6.13. The equity flow = the equityholder dividend – the capital contribution. The implied equity flow at time t=½ is –$104.87, so the capital contribution at time t=½ is $6.13 – (–$104.87) = $111.00.[24]

We examine the flow of funds in the first two periods. The original $412.50 plus the implied equity flow of $104.87 equals the equityholder funded capital of $523.50 at time t=½. The pre-tax investment income on the original equityholder funded capital was $16.49, and the dividend to the equityholders was only $6.13. The difference of $16.49 – $6.13 = $10.36 is transferred to fund the insurance policy.[25]

In the second period <<** Neeza to fill in **>>

SCENARIO B: FAIR VALUE PREMIUM

The policy premium of $1,058.27 covers the acquisition expenses of $252.89 and the present value at the pre-tax investment yield of future loss and expense costs of $805.38. The left hand side of the graphic shows the accumulated premium paying the insurance costs of each period. The remainder of the accumulated premium in each period covers the loss and expense liabilities.

The right-hand side of the graphic shows the flow of equityholder funds. The equityholders contribute $374.62 at time t=0 to support the policy.[26]

The pre-tax semi-annual investment yield on equityholders' funds is $374.62 × 4.0% = $14.98. The equityholders receive the full pre-tax investment yield; the federal income taxes on this investment yield are included in the policy costs paid by the policyholders. The internal rate of return is the pre-tax investment yield, so the $14.98 is paid as a shareholders dividend at time t=½. At that date, the equityholders make a second capital contribution of $119.85 to support the loss reserves. The additional capital contribution minus the dividend equals the implied equity flow at time t=½:

$119.85 – $14.98 = $104.87.

24 The implied equity flow is the combination of the equityholder dividend and any other capital exchanges. It equals the dividend minus any capital contribution or the dividend plus any return of capital.

25 The schematic shows this as the full investment income on the policyholder funded capital of –$37.88 plus a portion of the capital itself: $1.51 + $8.86 = $10.37.

26 The change in the policy premium is not equal to the change in the initial equity flow. In scenario B, the policy premium is $1,058.27 – $1,000.00 = $58.27 greater than in scenario A, and the initial equity flow is $412.50 – $374.62 = $37.88 lower. The policy premium is a one-time payment. The capital contribution made by the equityholders is held for several years.

The two values – the policy premium and the equityholder funded capital – are not in the same units. The policy premium is measured in dollars; the equityholder provided capital is measured in dollar-years. When the equityholder provided capital is multiplied by the cost of holding capital, which is a percentage amount per year, the product is measured in dollars. See Feldblum and Thandi, [2002], "Federal Income Taxes and the Cost of Holding Capital," for further discussion.

Although the equityholders' funds are not needed to fund the insurance operations, the policy is not adequately priced. The equityholders are receiving the benchmark investment yield of 8.0% per annum, when the opportunity cost of equity capital is 12% per annum. The equityholders have assumed the risks of insurance operations. If the insurance enterprise is unprofitable, the equityholders fund the losses.[27]

[27] Some analysts differentiate between investment risks and underwriting risks. This may be useful in distinguishing systematic risks from diversifiable risks. We do not differentiate risks in this fashion. Whether the risks of insurance operations are underwriting risks or investment risks, the equityholders fund the losses.

SCENARIO C: ADEQUATE PREMIUMS

In Scenario C, the internal rate of return exactly equals the cost of equity capital. We divide the policyholder premium into two segments. One segment covers the policy costs, including the federal income taxes on the investment income on equityholder supplied funds. The other segment is the capital (or profit) supplied by the policyholders. This capital is transferred incrementally to the equityholders to fund the difference between the investment yield and the cost of equity capital.

Illustration: The policyholder funded capital (PFC) in the original premium is \$24.57. The remainder of the premium, or \$1,096.07 – \$24.57 = \$1,071.50, funds the policy costs. The policy costs equal the paid amounts of \$266.13 plus the present value of future costs (at the pre-tax investment yield) of \$805.38:

$$\$266.13 + \$805.38 = \$1,071.51.$$

At time t=½, the equityholders receive the pre-tax investment yield from their own funds, or \$350.05 × 4.0% = \$14.00. The equityholders require a return of 5.83% each half year. The remaining 1.83% comes from the policyholder funded capital.[28]

At time t=½, the equityholders contribute additional capital of \$125.27 to support the loss reserve. The implied equity flow at time t=½ is

$$\$125.27 - \$16.37 = \$104.90.$$

We continue in this fashion through all valuation periods. By the end of the third year, the full policyholder funded capital of \$24.75 has been transferred to the equityholders.

[28] The graphic portrays this as the investment income on the policyholder funded capital plus a portion of the capital itself: \$0.98 + \$5.41= \$6.39 = 1.83% × \$350.05 = \$6.41.

Accounting System Graphics

The accompanying graphics show after-tax net income and capital amounts for the six accounting systems discussed in this paper. The accounting systems are grouped into three pairs:

- GAAP and statutory accounting
- NPV and IRR accounting systems
- Fair value and tax accounting

The accounting system graphics apply to the illustration in the text. The comments in the text relating to common elements for each pair of accounting systems refers to the expected income recognition pattern for property-casualty insurance products, not to the illustration in the text. We have deliberately chosen a somewhat unusual expense and loss pattern for the illustration, to better illustrate the computation of the deferred tax assets and the equity flows. The patterns in the graphics are close to the generalizations in the text, though they are not identical.

Appendices

Appendix A to this paper shows the equivalence of the net present value of the EVA's over the lifetime of the project under any accounting system. The demonstration applies to a consistent accounting system as defined in this paper; see the comments above about direct charge and credit to surplus under statutory accounting and tax-exempt income under tax accounting. The reasoning in the appendix follows from the definition of the economic value added and the relationships among income and capital between accounting systems. The intuition for our treatment of accounting systems is provided in the text of the paper; the appendix provides a formal mathematical proof.

Appendix B to this paper, contributed by Dr Ernesto Schirmacher, provides an alternative perspective for viewing the implied equity flows under any accounting system. In the text of the paper, we have provided separate analyses of each accounting system, along with the unique characteristics of each of them. Appendix B shows that all the accounting systems can be viewed from the same perspective, but the rate of return varies from one system to another in each accounting period.

Financial Pricing Models for Property-Casualty Insurance Products: Reserve Valuation Rates

Sholom Feldblum, FCAS, FSA, MAAA, and
Neeza Thandi, FCAS, MAAA

Financial Pricing Models for Property-Casualty Insurance Products: Reserve Valuation Rates

ABSTRACT

This paper analyzes the relationship between the reserve valuation rate and the indicated premium rate. The reserve valuation rate affects the capital embedded in the reserves. Full value loss reserves contain much embedded capital, whereas fair value reserves contain little (if any) embedded capital.

The amount of embedded capital in the reported loss reserves affects the return on capital. If the insurer prices its policies to achieve a target return on capital, the amount of embedded capital affects the premium rates.

This paper discusses the underlying concepts, and it presents the intuition for the analysis. A companion paper, "Federal Income Taxes and the Cost of Holding Capital," extends the analysis to include the effects of federal income taxes and the cost of holding capital. The illustration in this paper is carried over to the companion paper.

TERMINOLOGY

The reserve valuation rate in life insurance is the discount rate used to value the policy reserves. The discount rate is constrained by statutory regulation. For the maximum permitted reserve valuation rate, the NAIC Standard Valuation Law (1990) uses a dynamic standard, based on investment grade corporate bond yields minus a specified margin. The exact valuation rate depends on the characteristics of the insurance product, such as the reserve duration, surrender charges, and market value adjustments.[1]

For property-casualty business, statutory accounting requires full value reserves. This was economically beneficial before 1986, since it helped property-casualty insurance companies defer federal income taxes on their underwriting operations.

This deferment of federal income taxes ended with the Tax Reform Act of 1986, which set a 60 moving average of the federal mid-term rate as the valuation rate for tax basis reserves. After 1986, full value statutory reserves have been justified as a means of providing a risk margin in adverse scenarios, thereby helping maintain the solvency of companies.

---

1 The codification of statutory accounting has taken a step in the same direction by setting a dynamic formula for the maximum interest rate permissible for non-tabular discounting; see SSAP 65, "Property and Casualty Contracts," paragraph 12.

With the advent of risk-based capital requirements in 1992, statutory solvency monitoring uses discounted reserves. The RBC formula uses a 5% loss reserve valuation rate coupled with the IRS loss payment patterns by line of business. The reserving risk charge in the RBC formula is expressed as an explicit capital requirement, not as a component of loss reserves.[2]

LOSS RESERVE DISCOUNTING

Property-casualty statutory accounting requires full value (undiscounted) reserves, except in certain limited circumstances:

1. Tabular reserve discounts are permitted on the indemnity portions workers' compensation long term disability claims (pension cases) and on long term disability claims written on accident and health insurance policies. These are annuity claims on impaired lives. Just as they are discounted on the life insurance statutory blank, they are discounted on the fire and casualty blank, whether the policies are written by life insurers, health insurers, or workers' compensation insurers. Tabular discounts are not permitted for medical benefits or for loss adjustment expenses, even if these benefits are paid on the same claims.
2. Reserve discounts are permitted for certain monoline (primarily single-state) medical malpractice writers. This regulation was designed to help privately organized "doctors' mutuals" write medical malpractice coverage without having to raise additional capital.
3. Reserve discounts may be specifically allowed by the insurance commissioner of the domiciliary state. These discounts are intended to enable a domestic company to continue operating even with low statutory surplus.

These three instances are *explicit* reserve discounts. Only the first of these (workers' compensation tabular discounts) is relevant to general property-casualty pricing models.

This paper deals with *implicit* reserve discounts. We differentiate among three items:

4. An *unintended* reserve deficiency stems from miscalculation of the indicated reserves. Sometimes this reflects poor actuarial judgment; sometimes this reflects unforeseeable legal, social, or economic developments. For instance, the surge in asbestos claims in the late 1990's and early 2000's was an unforeseen social and legal phenomenon which raised reserve indications for general liability.
5. An *intended* reserve deficiency is a conscious management decision to hold less than full value reserves to improve the reported surplus of the company. It is a company-wide surplus management decision, not a line of business pricing decision. The cost of an intended reserve deficiency is the increased present value of federal income tax liabilities.
6. An *implicit reserve discount* uses the present value of loss reserves, where the discount rate is the reserve valuation rate. The objective is income optimization, stemming from the reduced cost of holding capital. This is partially offset by the increased present value of federal income tax liabilities.

---

[2] On the gradual change of property-casualty reserve valuation from a full value basis to a fair value basis, see Feldblum [1994: LRD].

We differentiate between an intended reserve deficiency and an implicit reserve discount:

Intended reserve deficiency: A company in weak financial condition may lower its carried loss reserves to show greater capital and surplus funds. If the company makes no other change in its operations, its return on invested capital is unchanged. The invested capital is simply moved from the loss reserves to policyholders' surplus.

An *implicit reserve discount* removes equityholder provided capital from the loss reserves and from policyholders' surplus. These funds may be

7. returned to the equityholders by means of stockholder dividends or repurchase of shares (or by policyholder dividends in a mutual insurance company), or
8. used for other purposes, such as to write more premium in profitable lines of business, to expand into other geographic areas, or to engage in other activities, such as financial services.

The implicit reserve discount reduces the implied equity flows and raises the return on invested capital. A major responsibility of corporate management is to make optimal use of investors' capital. From this perspective, the misuse of capital might be viewed as a dereliction of duty. Allowing equityholders' capital to sit idly in reserves and incur double taxation might be viewed as poor capital management.[3]

The received wisdom in the insurance industry is that greater reserve adequacy is better, since consumers seek insurance companies that are financially strong, and companies with more adequate reserves are less likely to fail. For well-managed and financially stable companies, this reasoning is not always true. Higher reserve adequacy may indicate poor capital management, a lower return on capital, and higher policy premiums. If the increased risk of insolvency is not material, many consumers would prefer lower premiums.

For a given premium rate, a higher anticipated reserve adequacy causes a lower return on capital. In a line of business where peer companies are holding partially discounted reserves, an insurer with full value reserves may be at an economic disadvantage.

CASH FLOWS VS EARNINGS

The anticipated reserve adequacy is an accounting phenomenon; it does not affect the underwriting cash flows. It *does* affect the federal income tax cash flows, the assets required to support the insurance operations, the capital requirements, and the implied equity flows.

The anticipated reserve adequacy would have little effect on product pricing in a non-regulated industry. It has been ignored by some financial analysts developing insurance pricing models, who have focused more on the company's cash flows than on implied equity flows. The resultant rate

3 A risk of implicit reserve discounts is rating agencies might require greater statutory surplus. If management has indeed improved the company's return on capital and put the excess capital to better use, their actions may be viewed favorably by rating agencies.

indications are biased downwards, since they do not take into account the full cost of holding capital. For instance, Myers and Cohn [1978] use fully discounted loss reserves. They explicitly admit the inconsistency with statutory accounting (p. 67, footnote 1):

> *This view of policy reserves differs from the usual statutory insurance accounting view of posting full nominal or undiscounted reserves for losses and expenses.*

This can cause the Myers/Cohn model to understate the rate indications.[4]

Insurance rate filings assume that loss reserves are held at undiscounted values; any other assumption would contravene statutory requirements. A regulatory pricing model must assume full value reserves. The current use of the Myers/Cohn model in Massachusetts, with no adjustment for the cost of capital embedded in undiscounted loss reserves or gross unearned premium reserves, is inconsistent with regulatory requirements.

If one's peer companies are holding less than full value reserves, an assumption of full value reserves would produce non-competitive rates. A pricing model for a competitive insurance market should use the level of reserve adequacy expected for the block of business.

Cost of Holding Capital

The cost of holding capital is the difference between the cost of equity capital and the after-tax investment yield, adjusted for any additional taxes paid on the funds used to reimburse this cost. The illustration below uses a 10% investment yield and a 15% cost of equity capital, leading to a $15\% - (1 - 35\%) \times 10\% = 8.5\%$ per annum cost of holding capital, exclusive of additional taxes paid on the funds used to reimburse this cost. Each dollar of capital held by the company for a period of one year costs the equityholders 8.5¢.

The policyholders pay this cost through the profit margin in the policy premium. The policy premium is a pre-tax cash flow, and the cost of holding capital is an after-tax cost. The 8.5% after-tax cost is equivalent to an $8.5\% / (1 - 35\%) = 13.08\%$ addition to the policy premium. If the premium is paid at policy inception, the 13.08% must be discounted to the beginning of the year: $13.08\% / 1.100 = 11.89\%$.

---

4 For discussion of the Myers/Cohn pricing model, see Feldblum [Disc of D&D].

A dollar of capital embedded in reserves for five years to fund the statutory full value reserves is equivalent to an amount of $1 + 1/(1+i) + 1/(1+i)^2 + 1/(1+i)^3 + 1/(1+i)^4 = (1 - v^5)/d$ = \$4.17 held for one year.[5] The cost to the policyholder is \$4.17 × 11.89% = \$0.50.

At an investment yield of 10% per annum, embedding a dollar of extra capital in the loss reserves for five years costs the policyholder 50¢ in extra premium.

Conclusions

The illustrations in this chapter and the following chapter highlight the relationships among the valuation rate, the implied equity flows, the tax liability, the return on capital, and the indicated premium rate. We summarize the relationships and the pricing implications.

The reserve valuation rate is sometimes seen as an internal accounting matter, with no material effect on the company's cash flows or the indicated premium rate. It is not considered in traditional actuarial ratemaking, and it is sometimes neglected even in financial pricing models.

The true effect of the reserve valuation rate is too substantial for such cursory treatment.

- For the commercial casualty lines of business, the cost of holding capital is one of the largest costs in providing insurance coverage.
- Other insurance costs are needed to service the business. These costs include underwriting services, policy issuance, loss engineering, claims handling, and general home office expenses. The costs of holding full value reserves is a regulatory mandate.
- Most of the cost of holding capital goes to the IRS. The statutory requirements for full value reserves transfers funds from policyholders to the U.S. Treasury.[6]

The objective of insurance regulation is to safeguard the interests of policyholders, not to act as a collection agency for the IRS. The NAIC and the actuarial community would do well to streamline statutory accounting for the benefit of insurance consumers.

RESERVE DISCOUNTING

---

5 In this formula, *v* is the reciprocal of unity plus the interest rate, or $v = 1/(1+i)$; *v* is also called the present value factor and sometimes the discount factor. The variable *d* is the discount rate, defined as $d = i/(1+i)$. If *i* is the interest rate paid in arrears, *d* is the corresponding rate paid in advance.

6 The federal income tax effects are analyzed in the companion paper, Feldblum and Thandi [2002] "Federal Income Taxes and the Cost of Holding Capital."

Implicit discounting of loss reserves reduces the capital requirements in two ways: there is less capital embedded in the reserves, and the RBC reserving risk charge is reduced. If the reserving risk charge is 10% of held reserves, each dollar of implicit reserve discounting reduces the capital requirements by $1.10.[7]

Illustration: With a 10% investment yield and a 15% cost of equity capital, the after-tax cost of holding capital is 8.5%. Including the risk-based capital reserving risk charge raises this to about $1.10 × 8.5% = $0.0935. The reduction in the policyholder premium stemming from a dollar of implicit discount is

$$[\$0.0935/(1-35\%)] / 1.1 = \$0.1308.$$

If the reserve is held for five years, the reduction in the policyholder premium stemming from a dollar of implicit discount is

$$[\$0.0935/(1-35\%)] \times 4.17 = \$0.5453.$$

FEDERAL INCOME TAXES

Implicit discounting of loss reserves speeds up the incurral of federal income tax liabilities, though it does not change the nominal tax liability over the lifetime of the claims. The cost of the faster incidence of the federal income tax liability is the investment income lost on the tax payment that is made too early.

Illustration: The ABC Insurance Co. incurs a loss of $100,000 on December 30, 20XX.

- Scenario A: the loss is reported quickly and recognized in the 20XX Annual Statement.
- Scenario B: the loss is reported several months later, and it is not recognized until the 20XX+1 Annual Statement.

Suppose the IRS loss reserve discount factor for accident year 20XX as of December 31, 20XX, is 90%. The pre-tax investment yield is 10% per annum. In both scenarios, ABC ultimately receives an offset of $100,000 to taxable income. In scenario B, the offset is received one year later than in scenario A. We calculate the increased tax cost of the later recognition of the loss.

The offset to taxable income from the early recognition of the loss in 20XX is 90% × $100,000 = $90,000. The reduction in the tax liability is $90,000 × 35% = $31,500.

The pre-tax investment yield is 10% per annum, so the after-tax investment yield is 10% × (1 − 35%) = 6.5% per annum. The after-tax investment income on the $31,500 of tax refund held for one year is $31,500 × 6.5% = $2,047.50.

7 The 10% risk-based capital charge is a lower bound; for most scenarios, the charge is about 25%.

The recognition of the loss one year earlier causes a cash gain equal to 2% of the loss. The total profit margin from underwriting and investment income in the insurance industry is about 7% of premium or about 9% of losses; the precise figure varies by company, by line of business, and by year. A one year deferral of loss recognition reduces the profit margin from 9% of losses to 7% of losses, for a 22% reduction. This stems solely from the tax effect.[8]

The net gain from implicit reserve discounting is the gain from freeing up capital minus the cost of deferring the recognition of losses. Using the figures above, a rough calculation gives 9.35% – 2.05% = 7.30%. This figure is correct only if the company has other uses for the freed-up capital that yield 15% per annum. If the released capital languishes idly in surplus, there is a net dollar cost of 2.05%.[9]

We have not yet considered the deferred tax asset resulting from IRS loss reserve discounting. In this example, the deferred tax asset from earlier recognition of the loss is \$100,000 × (1 – 90%) × 35% = \$3,500. Only a portion of this deferred tax asset is recognized on the statutory balance sheet. The portion depends on the payout pattern of this loss and on the IRS loss reserve discount factors for the line of business; the calculation procedure is shown in Appendix A of Feldblum and Thandi [2002], "Modeling the Equity Flows." As a rough estimate, the admitted DTA may be about \$1,000. This is 1% of the losses. If the after-tax cost of holding capital is 8.5%, the value of this DTA is \$85. This is 0.085% of the loss.

THEORY AND INTUITION

We use a heuristic example to show the effects of anticipated reserve adequacy. The example uses a one-day policy to avoid the complications of the IRS revenue offset provision and the capital embedded in the gross unearned premium reserve.

An company writes a one-day insurance policy on December 31, 20XX, for a premium of \$1,000. A loss occurs on that day, and it will be paid for \$1,000 on December 31, 20XX+3.

1. The pre-tax investment yield is 10% per annum.
2. The cost of equity capital is 15% per annum.
3. The tax rate is 35% on all income.
4. The required surplus capital is 20% of held loss reserves.
5. Acquisition expenses are \$170, paid on December 31, 20XX.

---

8 See also Feldblum and Schirmacher [2002: Reinsurance Pricing], who show the federal income tax effects of finite reinsurance, and Feldblum [2002: The Pricing of Commutations], who discusses the federal income tax effects stemming from claim commutations.

9 Even if the company has other uses for the capital, not all the benefit can be realized. If a company has less adequate reserves, rating agencies may require higher surplus. Perhaps half to two thirds of the capital is truly freed up; the remainder sits idly in surplus. The exact amount depends on the circumstances of each case.

To simplify the example, we assume that the IRS discounted reserves match the discounted value of the loss shown here.[10]

We determine the internal rate of return and the net present value at reserve valuation rates of 0% (current statutory accounting) and of 10% (fair value accounting). Full value statutory accounting requires more capital to be contributed by equityholders. Both the NPV and the IRR of the implied equity flows are lower if full value reserves are held than if fair value (discounted) reserves are held. We then re-price the policy at these two valuation rates such that the internal rate of return equals the cost of equity capital. The indicated premium rates are higher if statutory full value reserves are held than if fair value reserves are held.

The policyholder premium is a pre-tax figure; the NPV and the IRR are after-tax measures. The reserve valuation rate has a larger effect on the indicated premium than a cursory examination of the NPV might show.

ILLUSTRATION

The illustration speaks of a reserve valuation rate.

- At a 0% reserve valuation rate, the company holds full value reserves.
- At a 10% reserve valuation rate, the company holds fully discounted reserves. The anticipated reserve adequacy at policy inception is $1/1.100^3$ = 75.13% in this example.

Casualty actuaries speak of the level of reserve adequacy or the amount of implicit discount. If "i" is the reserve valuation rate and "n" is the average number of years between loss occurrence and loss payment, the anticipated reserve adequacy equals $1/(1+i)^n$.

Illustration: In the heuristic example above, "i" = 10% and "n" = 3. If losses are paid 3 years after they occur (on average) and the reserve valuation rate is 10% per annum, the anticipated reserve adequacy = $1/(1.10)^3$ = 75.13%. At a 10% reserve valuation rate, the level of reserve adequate is 90.91% if the loss will be paid in 1 year and 75.13% if the loss will be paid in three years.

PRICING AT A 0% VALUATION RATE

The illustrations in the companion papers use full value reserves, semi-annual valuations, and both acquisition and maintenance expenses. The focus in the present illustration is on the reserve valuation rate. We use annual valuation periods, with implied equity flows at December 31 of each year.

- From the cash flows, reserve changes, and capital requirements, we determine the implied equity flows.

10 This assumption is reasonable. For a real block of business, the IRS discount factors (over the long-run) are relatively unbiased estimates of the actuarially correct discount factors; see Feldblum [2002: SchP] and Sarason *et al.* [2002]. The 16 year limit to the IRS loss reserve payout pattern is not material in most lines of business. We note the magnitude of the tax effects on the rate indications later in this chapter.

- From the implied equity flows, we determine the NPV and the IRR.
- We show the resultant NPV and IRR at a 0% reserve valuation rate and a 10% reserve valuation rate.
- To actually price the policy, we solve for the premium rate that generates an NPV of zero or an IRR equal to the cost of equity capital. We show the pricing results in the exhibits below.

CASH FLOWS, DEFERRED TAX ASSET, AND EQUITY FLOWS

On December 31, 20XX, the gross premium is \$1,000 and the acquisition expenses are \$170. The loss reserves are \$1,000, and the required surplus is \$200.

Since this is a one day policy, the unearned premium reserve is \$0 at the end of the day. There is no tax effect from revenue offset, and there is no associated deferred tax asset.[11]

We assume that the IRS discount rate is also 10% per annum, and the IRS loss payment pattern corresponds to the actual loss payment pattern in this example. The discounted reserves for tax purposes are $\$1{,}000 / 1.10^3 = \$751.31$. The taxable underwriting income is

$$premium - expenses - discounted\ losses = \$1000.00 - \$170.00 - \$751.31 = \$78.69.$$

The tax liability is $\$78.69 \times 35\% = \$27.54$.

The gross deferred tax asset is $35\% \times (\$1000 - \$751.31) = \$87.04$. Statutory accounting recognizes the portion of the gross deferred tax asset which reverses within 12 months. Since no losses are paid in the coming 12 months, this amount equals the tax rate times the undiscounted loss reserves times the change in the IRS loss reserve discount factor from the current valuation date to the valuation date 12 months from now.

$$35\% \times \$1000 \times (1/1.100^2 - 1/1.100^3) = 35\% \times \$1000 \times (82.64\% - 75.13\%) = \$26.30.$$

The required surplus on December 31, 20XX, is $\$1000 \times 20\% = \$200$. The total required assets are \$1000 of loss reserves + \$200 of required surplus = \$1,200. The assets held by the insurance company are

	\$1,000 of premium
–	\$170 of acquisition expenses
–	\$27.54 of federal income tax payment
+	\$26.30 of deferred tax asset
=	\$828.76

The implied equity flow on December 31, 20XX, is a capital contribution of \$1,200 – \$828.76 = \$371.24. The investable assets are \$1,200 – \$26.30 = \$1,173.70.

[11] None of the conclusions here depend on the one-day policy term. In fact, the gross unearned premium reserves cause additional capital to be tied up in statutory reserves, magnifying the effects discussed in the text.

YEAR 20XX+1

- The investment income during 20XX+1 is \$1173.70 × 10% = \$117.37.
- The tax on the investment income is \$117.37 × 35% = \$41.08.
- The tax on underwriting income is 35% × (\$751.31 – \$826.45) = –\$26.30.
- The combined tax liability for 20XX+1 is \$41.08 – \$26.30 = \$14.78.

The deferred tax asset from December 31, 20XX, is eliminated at December 31, 20XX+1. A new deferred tax asset is set up on that day for 35% × \$1000 × (90.91% – 82.64%) = \$28.93. The change in the deferred tax asset is \$28.93 – \$26.30 = \$2.63.[12]

The net income in 20XX+1 equals

  \$117.37 of investment income
– \$14.78 of federal income tax payment
+ \$2.63 of change in the deferred tax asset
= \$105.22

The required surplus remains \$200 during 20XX+1. The implied equity flow equals the net income minus the change in capital. The implied equity flow on December 31, 20XX+1, is a payment to equityholders of \$105.22.

IRR AND NPV CALCULATIONS

We repeat this analysis for 20XX+2 and 20XX+3, as shown in Exhibit ??. The implied equity flows at the four valuation dates are shown below:

Valuation date	12/31/20XX	12/31/20XX+1	12/31/20XX+2	12/31/20XX+3
Implied equity flow	–\$371.24	+\$105.22	+\$107.94	+\$275.93

The internal rate of return on these implied equity flows is the solution to the equation

$$-\$371.24 + \$105.22/(1+x) + \$107.94/(1+x)^2 + \$275.93/(1+x)^3 = 0.$$

The solution is x = 12.68%. The cost of equity capital is 15.0% per annum. The internal rate of return is lower than the cost of equity capital, and the policy is not profitable.

At a 15% cost of equity capital, the net present value of this policy is

$$-\$371.24 + \$105.22/(1.15) + \$107.94/(1.15)^2 + \$275.93/(1.15)^3 = -\$16.71.$$

---

12 The deferred tax asset increases in this illustration because there are no interim loss payments. When losses are paid gradually over the years, the deferred tax asset decreases steadily to zero.

The NPV equals the economic value added at policy inception under an NPV accounting system. This is also the present value of the total EVA under any accounting system, if the EVA's are discounted at the cost of equity capital. The premium must be increased by at least $16.71 for the insurer to break even on the policy.[13]

PRICING AT 10% VALUATION RATE

A 0% valuation rate represents full value reserves. A 10% valuation rate represents reserves discounted at the investment yield. These are the reserves that would be held in a fair value accounting system, except that here we assume the discounting is implicit. The implicit discounting raises the present value of the federal income taxes and reduces the benefits of the deferred tax asset.

The polar cases of a 0% valuation rate and a 10% valuation rate highlight the pricing and profitability effects of the reserve valuation rate. The exhibits at the end of this chapter also show the results for a partial discount at a 5% valuation rate. An implicit discount midway between full value reserves and fair value reserves better reflects the practice in the long-tailed lines of business and probably represents a better use of equityholders' capital.

The implied equity flows change in several ways when the valuation rate changes to 10%.

---

[13] A premium increase just equal to the EVA is generally insufficient, for two reasons:

- Increasing the premium increases the variable expense costs, such as agents' commissions.
- Increasing the premium generally increases the capital requirements.

In theory, higher premium rates should lead to reduced risk-based capital requirements. In practice, the written premium risk charge is a direct function of the premium rates, not an inverse function.

1. *The assets needed to back the loss reserves change.* At a 0% valuation rate, the December 31, 20XX, loss reserves are \$1,000. At a 10% valuation rate, the December 31, 20XX, loss reserves are \$751.31 . The change in the implied equity flow stemming from the difference in the loss reserves is \$1,000 – \$751.31 = \$248.69.[14] (The full change in the implied equity flow incorporates other items as well.)
2. *The risk-based capital requirements change.* The required capital is 20% × \$1,000 = \$200 if a 0% valuation rate is used and 20% × \$751.31 = \$150.26 if a 10% valuation rate is used. The difference in the implied equity flow on December 31, 20XX, is \$200.00 – \$150.26 = \$49.74.[15]

In theory, the company's capital requirements should depend on indicated reserves, not held reserves. If the company holds less assets to back the loss reserves with a 10% valuation rate, it should hold more capital to offset the increased insolvency risk. In practice, the RBC formula sets the capital requirements as a function of held reserves. It makes no attempt to quantify the adequacy of these reserves.[16]

---

[14] The assets needed to fund the unearned premium reserves on annual term policies show a similar effect. After adjusting for the deferred tax asset, the difference between gross and net unearned premium reserves is 65% of the pre-paid acquisition costs on the unexpired portion of the policies.

[15] There is no corresponding effect on the unearned premium reserves, since there is no risk-based capital charge on the unearned premium reserves.

[16] The opposite is true for explicit reserve discounts. If the 10% reserve valuation rate is make explicit, the amount of the discount is removed from statutory surplus and the risk-based capital reserving risk charge is based on the undiscounted loss reserves.

The rating agencies do attempt to estimate the required reserves, and they may adjust the capital requirements for the difference between their estimates of required reserves and the company's held reserves. We do not attempt to model the effects of loss reserve valuation rates on rating agency capital requirements. This is an important consideration to keep in mind, but there is too much uncertainty to formulate a fixed pricing procedure.[17]

3. The timing of the federal income tax liabilities changes if the discount on the loss reserves is left implicit and not disclosed in the statutory Annual Statement. If the loss reserve discount is disclosed in the Annual Statement, the statement loss reserves are grossed up for the amount of the discount and there is no change in the IRS discounted reserves.[18]

The total dollar amount of taxes paid does not depend on the valuation rate. However, a higher valuation rate causes the tax liabilities to be incurred earlier, leading to a loss of investment income and a higher present value of the tax cash flows.

- With a 0% reserve valuation rate, (i) the IRS discounted loss reserves on December 31, 20XX, are 75.13% × $1,000 = $751.31, (ii) the IRS discounted loss reserves on December 31, 20XX+1, are 82.64% × $1,000 = $826.45, and (iii) the tax-basis incurred losses in 20XX+1 are $826.45 – $751.31 = $75.13.
- With a 10% loss reserve valuation rate, (i) the IRS discounted loss reserves on December 31, 20XX, are 75.13% × $751.31 = $564.47, (ii) the IRS discounted loss reserves on December 31, 20XX+1, are 82.64% × $826.45 = $683.02, and (iii) the tax-basis incurred losses in 20XX+1 are $683.02 – $564.47 = $118.54.[19]

Because the interest discount is not disclosed, the loss reserves are "doubly discounted" for computing taxable income. If the greater loss reserve discount were entirely offset by a deferred tax asset, the increased tax payment would *not* change the implied equity flow on December 31, 20XX. The only effect would be a decrease in the investment income received in 20XX+1, since the deferred tax asset is not an investable asset.

There are two reasons why the greater loss reserve discount is not entirely offset by a greater deferred tax asset.

✓ The change in the federal income tax liability is a multi-year effect. Statutory accounting admits only the portion of the deferred tax asset that will reverse within 12 months.

---

17 In practice, the company's chief actuary meets periodically with A. B. Best's and with Standard & Poor's to review the company's ratings and financial condition. These meetings generally give a good sense of how the rating agencies view the company's capital management stance. The information gained at these meetings is a necessary component for determining the optimal reserve valuation rate, even though we can not easily provide a generic formula to incorporate rating agency views.

18 The one constraint is the IRS discounted loss reserves may not be greater than the Annual Statement loss reserves (see Feldblum [2002: Schedule P]). This constraint affects the Schedule P "prior years" row for workers' compensation and long term disability insurance.

19 This illustration uses a valuation rate equal to the IRS loss reserve discount rate. This would be rare in the property-casualty insurance industry. The reserve valuation rate would generally be lower than the IRS loss reserve discount rate.

✓ Since the reserve discount is left implicit, the company must set up the deferred tax asset *as if the reserves were held at full value.* At a higher reserve valuation rate, the deferred tax asset decreases; it does not increase.

Illustration: The full deferred tax asset for a 10% valuation rate ought to be 35% × ($1,000 – $564.47) = $152.44. The company can not put up this deferred tax asset without acknowledging that its loss reserves are under-stated. Since the company holds a loss reserve of $751.31, the full deferred tax asset is the tax rate 35% × ($751.31 – $751.31 × 0.75131) = $65.40.

The effect on the statutory deferred tax asset is similar. The full statutory deferred tax asset on December 31, 20XX, ought to be 35% × ($826.45 × 0.82645 – $751.31 × 0.75131) = $41.49. The company can not put up this deferred tax asset without acknowledging that its loss reserves are under-stated. Since the company holds a loss reserve of $751.31 with a corresponding tax-basis reserve of $564.47, the deferred tax asset that can be admitted on the statutory balance sheet is 35% × [($751.31 – $564.47) – ($751.31 – $683.01)] = $19.76, or 35% × ($564.47/0.90909 – $564.47) = $19.76.

CASH FLOWS AND IMPLIED EQUITY FLOWS

To show the magnitude of these effects, we compute the cash flows, the accounting entries, and the implied equity flows for a 10% valuation rate.

On December 31, 20XX, the premium is $1,000 and the acquisition expenses are $170. The loss reserves are $1,000 / 1.10^3 = $751.31, and the required surplus is $150.26.[20]

The IRS discounted reserves are $751.31 / 1.10^3 = $564.47. The loss reserves are doubly discounted: once (implicitly) by the company and a second time (explicitly) by the IRS. The taxable underwriting income is

premium – expenses – discounted losses = $1000.00 – $170.00 – $564.47 = $265.53.

The tax payment is $265.53 × 35% = $92.94. The deferred tax asset is 35% × ($564.47 × 1.100 – $564.47) = $19.76, since the company records $751.31 as its statutory loss reserve.

The total required assets on December 31, 20XX, equal $751.31 of loss reserves + $150.26 of required surplus = $901.57. The assets held by the insurance company are

  $1,000 of written premium
– $170 of acquisition expenses
– $92.94 of federal income tax payment

[20] The reserve valuation rate has no effect on the unearned premium reserve, the revenue offset effects, or the deferred tax asset from revenue offset. In any case, we are using a one-day policy in this illustration, so this subject is moot; there is no unearned premium reserve.

+ $19.76 of deferred tax asset
= $756.82 of held assets

The implied equity flow on December 31, 20XX, is a capital contribution of $901.57 – $756.82 = $144.75. This is only $144.75/$371.24 = 38.99% of the capital contribution needed with a 0% valuation rate.

YEAR 20XX+1

- The investable assets are $901.57 – $19.76 = $881.81.
- Investment income during 20XX+1 is $881.81 × 10% = $88.18.
- The tax on the investment income is $88.18 × 35% = $30.86.
- The tax on underwriting income is 35% × ($751.31 × 0.75131 – $826.45 × 0.82645) = $-41.49.
- The deferred tax asset on December 31, 20XX+1, is 35% × ($683.01 × 1.100 – $683.01) = $23.91, since the tax reserve at this date is $683.01.

The total required assets on December 31, 20XX+1, equal loss reserves of $826.45 + required surplus of 20% × 826.45 = $165.29, for a total of $991.74. The assets held by the insurance company are

$901.57 of assets at December 31, 20XX
+ $88.18 of investment income
– $30.86 of federal income tax payment on investment income
+ $41.49 of federal income tax payment (a tax refund) on underwriting income
+ $23.91 – $19.76 of change in the deferred tax asset
= $1,004.53 of held assets

The implied equity flow on December 31, 20XX, is a capital distribution of $1004.53 – $991.74 = $12.79.

EQUITY FLOWS AND INTERNAL RATE OF RETURN

We continue in this fashion for years 20XX+2 and 20XX+3, as shown in Exhibit ??. The implied equity flows are shown below.

Valuation date	12/31/20XX	12/31/20XX+1	12/31/20XX+2	12/31/20XX+3
Implied equity flow	–$144.76	+$12.80	+$18.96	+$191.76

The internal rate of return on these implied equity flows is the solution to the equation

$$-\$144.76 + \$12.80/(1+x) + \$18.96/(1+x)^2 + \$191.76/(1+x)^3 = 0.$$

The solution is x = 16.93%. The cost of equity capital is 15.0% per annum. The internal rate of return is higher than the cost of equity capital, and the policy is profitable.

At a 15% cost of equity capital, the net present value of this policy is

$$-\$144.76 + \$12.80/(1.15) + \$18.96/(1.15)^2 + \$191.76/(1.15)^3 = +\$6.79.$$

This is also the economic value added under the NPV accounting system, which we use as the performance measure. Writing the policy increases the value of the company.

At a 0% valuation rate, the internal rate of return was 12.68%. The change in the reserve valuation rate, without any change in the underwriting cash flows, increases the IRR by 16.93 – 12.68 = 4.25 percentage points, turning a money losing policy into a profitable policy.

CONSUMER'S PERSPECTIVE

The illustration above assumes that the policyholder premium is fixed at $1000, and that insurers earn economic profits or suffer losses depending on the cost of providing coverage.

The actual U.S. insurance markets are highly competitive, with scores of competing firms in each region, low concentration ratios, no perceptible economies of scale, low barriers to entry, and no significant product differentiation. Rates of return in excess of the cost of capital do not persist. The lower cost of holding capital stemming from fair value reserves would reduce the policyholder premium.

Exhibits ?? and ?? show the policy priced to provide a 15% internal rate of return on the implied equity flows. At a 0% valuation rate, the indicated premium is $1025.70. At a 10% valuation rate, the indicated premium is $989.55. The change in the valuation rate reduces the indicated premium by 3.5%.

In both scenarios, the present value of expected losses and expenses is $921.31. The profit margin in the premium is $104.39, or 11.33% of the discounted net premium, for a 0% reserve valuation rate, and $68.23, or 7.41% of the discounted net premium, for a 10% reserve valuation rate. The change in the valuation rate reduces the profit margin by $36.16, or 34.64%.

SUMMARY

We summarize the effects of different valuation rates below; see Exhibit ?? for the calculations. The premium in each scenario is $1,000.

- *Full value loss reserves* (0% valuation rate) require the greatest capital investment ($317.24), but the net present value of the tax credit stemming from incurred losses is also greatest ($394.18).

- *Implicit discounting* (10% valuation rate) requires the least capital investment, though only if the capital requirements do not change because of the discounting. However, implicit discounting reduces the tax credit stemming from incurred losses to $381.91.
- *Explicit discounting* (10% valuation rate) requires an intermediate capital investment, and the full tax credit stemming from incurred losses is retained.

The table below shows the premium for the three scenarios in the list above if the policy is priced to yield a 15% internal rate of return.

	No Discounting	*Implicit Discounting*	*Explicit Discounting*
Premium	$1,025.70	$989.35	$981.03
NPV(FIT)	$66.76	$48.41	$38.31
NPV(tax credit on losses)	$394.18	$381.91	$399.18

Other Pricing Assumptions

The full analysis of the effect of the reserve valuation rate on the indicated premium rates requires careful consideration of four other components of policy pricing:

1. The target return on capital demanded by the company's equityholders.
2. The benchmark investment yield that the company expects to earn on its investable assets.
3. The expected reserve duration for the block of business being priced.
4. The effects of federal income taxes on both investment income and underwriting income.

This paper has dealt with the underlying concepts regarding the reserve valuation rate and its effects on policy pricing. The four issues listed above are complex, and they warrant more rigorous treatment than can be provided here. We deal with these four issues in detail in the companion paper, "Federal Income Taxes and the Cost of Holding Capital." The companion paper begins where this paper leaves off, and it provides a rigorous analysis of the interaction between the reserve valuation rate and the other pricing assumptions.

A Discussion of "Loss Estimates Using S-Curves: Environmental and Mass Tort Liabilities" by Bruce E. Ollodart

Kirk G. Fleming, FCAS, MAAA

A Discussion of "Loss Estimates Using S-Curves: Environmental and Mass Tort Liabilities"

by Kirk Fleming

Abstract: This paper is a discussion of Bruce Ollodart's 1997 Winter *Forum* paper on using S-Curves to model environmental and mass tort liabilities. To start, there is a brief summary of Ollodart's paper. Then I introduce a type of S-Curve known as the logistic curve. The logistic curve assumes a maximum number of claims so it eliminates at least one of the problems Ollodart mentions with the curves he discusses. Finally, I finish with some comments on the modeling process.

Bruce Ollodart in his paper "Loss Estimates using S-Curves: Environmental and Mass Tort Liabilities" proposes using S-Curves for the analysis of mass tort liabilities (e.g. environmental and asbestos). The appeal of these curves is that their shape matches how we have seen environmental claims emerge and be paid. The paid losses associated with these claims start out slowly, increase rapidly for a period of time and then finally slow down again. The cumulative payment pattern or cumulative reporting patterns follow this S pattern. Ollodart discusses potential candidates for S-Curves to use for this analysis as well as the strengths and weaknesses of these individual candidates.

In this paper, I discuss another type of S-curve called the logistic curve along with a justification for its use. This form of S-curve has a long history of use in economics and the social sciences. It can eliminate some of the shortcomings that Ollodart points out with some of the curves he uses.

A Quick Summary of Bruce Ollodart's Paper

Ollodart says that he has tested a number of curves as appropriate S-Curve models. One successful candidate is the power curve that he discusses in his paper. Another curve he discusses is the gamma curve.

The power curve has the following form:

$$Y = s\,(x-b)^P + c.$$

The variable Y represents the cumulative paid or reported losses, s is a scalar coefficient greater than zero; x is the year of projection (or year corresponding to the historical data), b represents the time at which the curve's inflection point occurs, P is an odd power between zero and one, and c is a constant representing the projected cumulative paid losses at time b. The power P is typically chosen from among the family of fractional powers 1/3, 1/5, 3/5, 1/7, 3/7, 5/7, 1/9, etc.

Ollodart mentions several problems with power functions. The curve increases without bound so the actuary must select a maximum runoff period. This selection has to be made possibly with little information to justify the selection. In practice the rate of change for the curve might be very low once you get out many years on the curve. It may not make a significant difference if you pick a 20-year runoff period or a 60-year runoff period. But it opens up an area for others to disagree with the methods used and the projections.

For example, the 1997 paper uses the power curve to project asbestos losses and the selection is a 20-year runoff period. Does it make a difference if we use 20 years or 60 years? Well depending upon the situation and the data, perhaps not. But if I am talking to the outside auditors, they might be reluctant to buy into the 20-year runoff period and if I am talking to my CFO he or she might not readily buy into a 60-years runoff. Either way, it's a conversation that we would rather not have after presenting results.

Another potential problem with the power curve is that it is very sensitive to the selection of the factors. To get around this problem Ollodart suggests restricting the parameter P to ten possible values. Given those ten different values, fit the curves to the data to get the remaining parameters in the curves. Then select the best curve from among those ten choices.

Logistic Curves

An excellent book to study, if people have the time, is Martin Braun's "Differential Equations and Their Applications". The book is an introduction to differential equations. Braun describes more and more complex and successful applications of differential equations in economics and the social sciences. We can use some of the techniques that Braun describes to construct a model of the claim payment process for mass torts.

Suppose the following story describes an environmental claim payment process. We are dealing with a type of pollution claim that was once covered under an occurrence policy but is no longer covered. Let us say the change in coverage took place in 1986. So there is the possibility of these pollution claims being reported from old policies but there will be no new claims of this type from policy years after 1986.

Initially, reporting and payments from these types of pollution claim were light. There was no incentive for insureds to report the claims and the regulatory agencies were not really pushing to get things cleaned up. Then some policyholders actually received some large settlements to clean up their pollution sites. As news of these settlements began to spread, other policyholders began to take notice. Regulators also began to take notice of the money available for clean up costs. As more and more insureds won settlements, it generated more and more reports of claims and demands from insureds for money for clean up. Eventually, the pool of reported and closed claims began to reach the maximum number of potential claims. As there were less and less claims that potentially could be reported, the rate of new reports slowed down. The same would be true with payments. As there were less and less claims that were yet to be settled, the rate of claims payments slowed down.

We can start to model this story for claim payments by saying that the population of cumulative claim payments is a function of time $Y(t)$. Suppose that N is the ultimate loss for all claims and suppose that c is a constant. As claims start to be settled successfully for large amounts, it causes more and more claims to be reported and settled. As we get closer to the ultimate loss dollars for all claims, the rate of new payments slows down. An initial value differential equation that describes this process is

$$dY/dt = c\,Y(N - Y) \quad \text{with} \quad Y(0) = 0. \qquad (1)$$

The solution to this problem is

$$Y(t) = N\exp(cNt) \,/\, (N - 1 + \exp(cNt)). \qquad (2)$$

There are a few desirable features about this solution. The graph of this solution is an S-curve. Also, it has a maximum value since that was one of the assumptions we started with. And finally, it is the result of modeling a process.

The equation (2) describes the logistics law of population growth. Braun points out that it was first introduced in 1837 by the Dutch mathematical-biologist Verhulst. It was an enhancement to the Malthusian law of population growth that had the unrealistic implication that populations grow to an infinite size. The logistic law assumes a maximum point for the population.

The logistic equation has been used to model many different growth patterns. One such pattern was the spread of technological innovations. Braun points out that an implication of the logistic equation is that growth speeds up to a point where the modeled population has reached its half way point. After that it slows down. At least for the spread of technological innovations, it seems that the actual data follows a pattern where the rate of growth slows down beyond the half way point. When describing the model for the spread of technological innovations, this could be explained by an enhancement to the story that allows for the impact of advertising in addition to word of mouth and an extra term in the differential equation. Let c' be a constant.

$$dY/dt = c\,Y(N - Y) + c'(N - Y) \quad Y(0) = 0. \qquad (3)$$

This extra term in the model says that when the number of people who have not heard of the technological innovation is large, there is a definite influence due to advertising.

The solution to this problem is

$$Y(t) = Nc'[\,exp(\,(c' + cN)\,t\,) - 1]\,/\,(c\,N + c'\,exp(\,(c' + cN)\,t\,)\,). \qquad (4)$$

The graph of this enhanced logistic equation has a maximum and will also be an S curve for appropriate choices of c and c'.

Using the Curve with the Original Data

The source of the original industry data in Ollodart's paper was confidential. Because the details behind the data are not available, I do not know what the claims process is and I was not able to construct an appropriate model to represent the claims process.

However, just to show that logistic curves can be fit to data and to give people a set of numbers they can use to check their work if they reproduce these equations, I refit the original data in Ollodart's paper with the enhanced logistics curve. I want to strongly emphasize that these calculations are not alternative projections of the results. One of my main points in this paper is that it is important to model a process rather than just fit a curve to data.

In practice when using a logistic curve, I have found the parameters *N*, *c* and *c'* using the Solver feature in Excel. I have the Solver minimized the sum of the squares of the actual points and the fitted points. Whether minimizing the sum of the squares of the actual points and the fitted points is the best function to minimize is up for discussion. Ollodart points out that it might be more appropriate to minimize a function that gives more weight to later data. That sounds like a good approach since later data is presumably more relevant.

The original asbestos data is shown on Exhibit 1 along with the power curve results and the results of the enhanced logistic curve. A graph is shown on Exhibit 2. Exhibit 3 has the original pollution data along the modeled power curve results and the enhanced logistic curve results. The first point in the pollution data looks too big to be the initial point. I assumed the first available point was actually year 6 as opposed to year 1 based on the annual change in losses. Exhibit 4 shows a graph of the results.

Curves With and Without Stories

In the spirit of provoking discussion, I will throw out the following thought – in some circles, data mining is considered a bad thing. Now I am hesitant to say that because I have friends who think data mining is a really good thing. And I suppose the explanation must be found in the way we each think about data mining.

There is a paper on the Chartered Financial Analyst syllabus called "Using Economic Models" written by Avery B. Shenfeld, the Senior Economist for CIBC Wood Gundy. In the paper he discusses different forecasting approaches and he discusses problems in the use of models. One of the potential problems that he discusses with the modeling process is data mining. By data mining he is referring to a process where the researcher will do multiple calculations with the data in order to get something that works. So in a sense, the modeler just lucks out in finding something that works on the past data but has no explanation for why it should work going forward.

When I first read the Ollodart paper, the process of continually fitting different S-curves to the data with the only justification being that the data looked like an S-curve struck me as being open to this type of criticism. The process described was modeling the data as opposed to modeling a process.

Some actuaries would argue that our field is threatened by other professionals who are just as qualified to do the same type of analysis as we do. We have to be careful to construct models with their appropriate inputs and solve for the implications of those models. Then we have to accept or reject the results of those models and assumptions based on our best judgment along with the input and insights of other experts. For those people who defend data mining, my guess is that they argue data mining produces a model that had previously gone unnoticed. Once the model is uncovered, the modeler would only use it if they understood how the model should work going forward.

As I wrote, it's a point for discussion.

Other Ideas

One of the reviewers of this paper asked, "I am curious, is there a reason that the S shaped curves do not 'work' for relatively shorter tail lines such as medical malpractice and workers compensation?" My answer to that is, "Who says they don't work?" I have not used S-Curves for development work because there are other accepted loss development models. There are papers that discuss using mathematical curve models for the development process such as Richard Sherman's useful and practical paper, "Extrapolating, Smoothing and Interpolating Development Factors". I would say the work is still to be done to see if there is a model of the claims process that justifies the use of an S-Curve on lines other than environmental.

Closing Comments

In corresponding with Bruce Ollodart about his paper, Bruce pointed out that it is important for others to take the basic ideas proposed by some and work to develop them. I certainly agree with that. All of the work that we do is building on things that others have done.

Given that, I would have to give thanks to my all my teachers. That would include Martin Braun for writing the book so that I could lift material directly from it. I also have to give thanks to all the people that I have worked with over the years and all the people that I work with now. Finally, thanks go to Bruce Ollodart and all casualty actuaries who have built and expanded the Casualty Actuarial Society so that we all have a profession to share.

Exhibit 1

Asbestos Indemnity and Expense
Cumulative Paid Loss
Original Data from Ollodart Paper
(000's)

CY	(1) Actual Calendar Yr Cumulative Paid Loss	(2) Annual Change In Losses	(3) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(4) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve	CY	(5) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(6) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve
1978	362	362		18,455	2003	3,624,338	8,777,269
1979	17,918	17,556	57,426	41,236	2004	3,693,242	9,596,461
1980	33,987	16,069	117,252	69,342	2005	3,758,660	10,377,967
1981	84,014	50,027	179,775	103,989	2006	3,821,037	11,108,040
1982	193,596	109,582	245,358	146,660	2007	3,880,734	11,776,818
1983	258,994	65,398	314,454	199,154	2008	3,938,045	12,378,529
1984	284,030	25,036	387,632	263,638	2009	3,993,215	12,911,201
1985	324,534	40,504	465,635	342,713	2010	4,046,451	13,376,037
1986	374,068	49,534	549,452	439,475	2011	4,097,929	13,776,626
1987	612,636	238,568	640,459	557,567	2012	4,147,800	14,118,133
1988	752,146	139,510	740,659	701,232	2013	4,196,194	14,406,601
1989	898,011	145,865	853,169	875,330	2014	4,243,225	14,648,374
1990	1,026,623	128,612	983,338	1,085,315	2015	4,288,996	14,849,691
1991	1,259,167	232,544	1,141,855	1,337,149	2016	4,333,593	15,016,410
1992	1,585,463	326,296	1,357,474	1,637,124	2017	4,377,097	15,153,856
1993	2,078,939	493,476	2,095,513	1,991,557	2018	4,419,578	15,266,747
1994	2,470,635	391,696	2,591,167	2,406,346	2019	4,461,100	15,359,187
1995	2,835,848	365,213	2,802,383	2,886,366	2020	4,501,720	15,434,691
1996			2,959,027	3,434,733	2021	4,541,489	15,496,237
1997			3,088,106	4,051,996	2022	4,580,456	15,546,320
1998			3,199,887	4,735,365	2023	4,618,663	15,587,021
1999			3,299,556	5,478,147	2024	4,656,149	15,620,061
2000			3,390,156	6,269,559	2025	4,692,951	15,646,857
2001			3,473,648	7,095,057	2026	4,729,102	15,668,574
2002			3,551,384	7,937,246	2027	4,764,633	15,686,165

Power Curve 7
Fulcrum Year 1993

Logistic Curve Paremeters

N	15,760,600
c	1.35E-08
c prime	1.05E-03

Asbestos Indemnity and Expense Cumulative Paid Loss Data from Ollodart's Paper

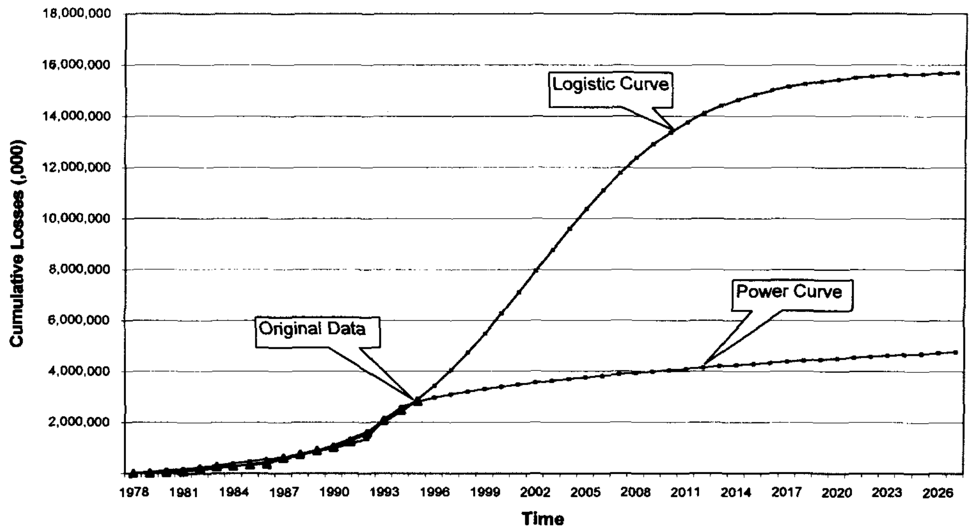


Exhibit 2

Exhibit 3

Pollution Indemnity and Expense
Cumulative Paid Loss
Original Data from Ollodart Paper
(000's)

CY	(1) Actual Calendar Yr Cumulative Paid Loss	(2) Annual Change In Losses	(3) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(4) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve	CY	(5) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(6) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve
1983	135,953	135,953		184,894	2006	5,922,432	6,020,415
1984	172,946	36,993	160,048	255,951	2007	6,072,103	6,077,768
1985	222,134	49,188	326,616	347,749	2008	6,217,219	6,121,693
1986	407,273	185,139	500,739	465,358	2009	6,358,232	6,155,202
1987	579,370	172,097	683,772	614,438	2010	6,495,522	6,180,686
1988	914,273	334,903	877,553	800,879	2011	6,629,411	6,200,022
1989	1,150,537	236,264	1,084,678	1,030,149	2012	6,760,177	6,214,668
1990	1,410,354	259,817	1,309,057	1,306,319	2013	6,888,061	6,225,746
1991	1,613,107	202,753	1,557,103	1,630,820	2014	7,013,275	6,234,117
1992	1,951,047	337,940	1,840,889	2,001,161	2015	7,136,002	6,240,438
1993	2,334,475	383,428	2,189,868	2,410,024	2016	7,256,409	6,245,208
1994	2,779,049	444,574	2,791,813	2,845,207	2017	7,374,642	6,248,806
1995	3,373,188	594,139	3,557,863	3,290,741	2018	7,490,832	6,251,519
1996			3,914,670	3,729,083	2019	7,605,098	6,253,565
1997			4,202,122	4,143,787	2020	7,717,546	6,255,107
1998			4,452,431	4,521,847	2021	7,828,275	6,256,268
1999			4,678,395	4,855,025	2022	7,937,373	6,257,144
2000			4,886,715	5,140,000	2023	8,044,921	6,257,804
2001			5,081,439	5,377,579	2024	8,150,993	6,258,301
2002			5,265,244	5,571,447	2025	8,255,658	6,258,675
2003			5,440,014	5,726,903	2026	8,358,979	6,258,957
2004			5,607,135	5,849,815	2027	8,461,015	6,259,170
2005			5,767,665	5,945,921			

Power Curve 9
Fulcrum Year 1995

Logistic Curve Paremeters

N	6,259,819
c	4.49E-08
c prime	1.93E-03

Pollution Indemnity and Expense Cumulative Paid loss Data from Ollodart's Paper

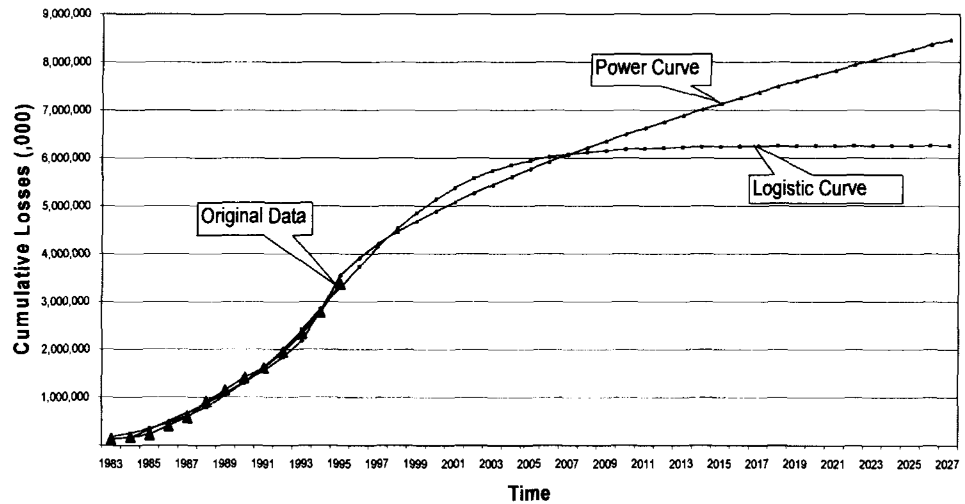


Exhibit 4

Bibliography

Braun, Martin 1993. Differential Equations and Their Applications: an Introduction to Applied Mathematics. New York: Springer-Verlag

Ollodart, Bruce C. 1997. "Loss Estimates Using S-Curves: Environmental and Mass Tort Liabilites," Casualty Actuarial Society Forum Winter 1997: 111-132.

Sherman, Richard E. 1984. "Extrapolating, Smoothing and Interpolating Development Factors," Proceedings of the Casualty Actuarial Society LXXI: 122-155.

Shenfeld, Avery B. 2001. "Using Economic Models," 2002 CFA Level III Candidate Readings Volume 1: 7-11.

Econometric Modeling of Insurance Frequency Trends: Which Model Should We Choose?

Amin Ussif, Ph.D.

Econometric Modeling of Insurance Frequency Trends: Which Model Should We Choose?

Amin Ussif[1], Ph.D.
223 6th Street
Fairview, New Jersey 07022
aussif@hotmail.com

Abstract

In policymaking and insurance rate setting process, understanding and managing claim frequency are crucial issues. Owing to the importance attached to the dynamics of claims frequency in insurance ratemaking and in implementing workplace safety measures, we intend to walk through the basic steps in the econometric modeling and forecasting of claims frequency. Data from the California Workers Compensation Institute (CWCI) are used in this study. Three competing models are developed with the goal of selecting a superior one amongst the three. All three specifications confirm the prior finding of the CWCI that economic activity is a significant determinant of workers compensation frequency. The conclusion is that the nonlinear models, (constant elasticity and the exponential or growth models) perform better than the linear model. Also, applying the likelihood ratio test and the F-test to the Actuarial models against the Econometric models, it is shown that considerable statistical gains can be achieved by using economic variables in estimating trends.

[1] I am grateful to Dan Corro and Greg Engl of NCCI, and Dr. Rashid Sumaila, Dr. Steve Morey and Betsy Fadali for their help.

Introduction

The goal in this paper is to investigate the effects of different functional form specification in modeling workers compensation claim frequency. While the economic theory of claims frequency is unambiguous, it is still not very clear how the frequency of claims filing is functionally related to the explanatory variables or covariates. One of the basic assumptions of linear regression is that the model is correctly specified thus making the choice of functional forms an important step in econometric modeling of claims frequency.

Understanding basic functional relationships is in fact very critical in the application of econometric modeling in practice. Forecasting is another very crucial aspect of insurance business, economics and finance. The health of the insurance industry depends on the accuracy of the forecasts. In setting premium rates, losses are forecast in advance and then rates are determined to cover claims when they occur.

This paper has a dual focus. First, it investigates the relationship between frequency and two key economic variables; employment and the unemployment rate. We use three different functional form specifications to test the various hypotheses about the impact of economic activity. The first functional form is the linear model, the second is the multiplicative model (a generalization of the Cobb-Douglas production function in economics) and the third is the exponential or growth model called the semi-log model in the econometrics literature.[2] The performance of these models is studied in order to objectively select the superior functional form based on statistics. The selection is based solely on information from the data and very little judgment is applied in order to maintain objectivity. Although the selection will be based on insample information, the real test of the quality of the models will be determined when we compare the predictive power of the models against experience outside the observation window. To our knowledge, this kind of study is the first ever that is conducted using the quarterly data from CWCI. Finally, we compare trend indications prior to credibility using the commonly employed models by actuaries, i.e. linear and exponential trend models. These actuarial trend models are special cases of the econometric counterparts developed in this paper.

---

[2] The second specification is the model developed by Kahley (2000) for the forecasting of frequency of claims for the California Workers Compensation Institute (CWCI).

The first section discusses the theoretical basis of the models. The economic theory of claims frequency filing is also presented and discussed. It then builds three mathematical models allowing for relevant nonlinearities. Several practical issues are also addressed. The final section discusses the empirical results and concludes the paper.

The Economic theory of Claims Frequency[3]

In general, it will be assumed that frequency is a function of employment, unemployment rate, and a trend. The basic hypothesis is that economic activity is a determinant of workers compensation claims frequency, i.e. increases in economic activity lead to increases in frequency holding other factors constant. It is also hypothesized that there are other ways that the economic environment affects the claim filing activity apart from the effects on the incidence of injuries. For example, the availability of jobs and the health of the labor market as reflected by changes in unemployment rate, plant closures, layoffs, etc., are potentially important causes of the incidence of claim filing (Kahley 2000). Another plausible postulate is that certain variables such as technology, safety initiatives from employers, etc impact the accident rates (see Ussif, 2002). For instance, technical progress and increases in workplace safety will a priori reduce incidence rates. We now specify various functional forms based on the above hypothesis.

Functional Form Specifications

Three alternative econometric models are considered in this section. All models have the same number of explanatory variables but differ only in their functional form specification. The first specification is linear in the variables while the others are nonlinear. The econometric models are given by the following equations

$$Y_t = \beta_1 + \beta_2 Emp_t + \beta_3\, Unemp_t + \beta_4 Time + \varepsilon_t \qquad (1)$$

$$Y_t = \beta_1 Emp_t^{\beta_2} Unemp_t^{\beta_3} \exp(\beta_4 Time + \varepsilon_t) \qquad (2)$$

$$Y_t = \exp(\beta_1 + \beta_2 Emp_t + \beta_3 Unemp_t + \beta_4 Time + \varepsilon_t) \quad (3)$$

---

[3] Frequency is defined as number of claims per earned premium. Please see Kahley for more about the data.

where exp(.) is the exponent function, *Emp* is the employment, *Unemp* is the unemployment rate and *Time* is the time trend dummy. Equation (2) is what is usually called the logarithmic-linear model while equation (3) is the semi-logarithmic model. The log-linear model has the advantage that the coefficients are the partial elasticities with respect to the independent variables. They simply tell us that a one percentage change in the independent variable will result in a certain percent change in the dependent variable. It is also clear that this model produces the average frequency growth rate frequency as the partial derivative of the dependent variable with respect to time. Equation three is the so-called exponential or growth model and its partial derivative with respect to time also gives the average frequency growth rate. These models extend the actuarial trend model to include economic covariates the unemployment rate and employment. By including these explanatory variables, the chances of capturing turning points may be greatly enhanced. To reiterate, these models may have some additional forecasting ability because of the information they used from the additional explanatory variables.

An important distinguishing feature of the models is that, the coefficient of the time trend variable in the linear model yields the absolute change in frequency per unit of time while the log-linear and the exponential give the percent change per unit of time. Hence, the nonlinear models have the additional advantage that they produce the equivalent of the actuarial trend estimates automatically.

In the application of econometric modeling to test refutable hypothesis and in forecasting, an important question is what makes a model "good"? To answer this question, we state a few criteria often used to help judge the "quality" of a model.

- Parsimony: - A mathematical model is a simplification of reality. It is not meant to capture all minor and random events but rather the essence of the phenomenon. All things being equal simpler models are preferred to unnecessarily large models. Simplicity in this context refers to the number of regressors and functional form.
- Theoretical consistency: - The coefficients in a model should have the right signs. A model may not be good if one or more of the estimated coefficients have the wrong signs. This has an important implication when using the model for purposes of forecasting.
- Goodness of fit: - A high adjusted R-square is good but this should not be overemphasized. Note that a model may not be good despite a high R-square if the estimated coefficients do not all have the

right signs. The main goal should not be to maximize the R-square but a good model with a high R-square is always welcome.

- Predictive power: - A good test of the validity of a model is comparison of its forecast with experience, its postsample predictive power. This also underscores the fact that a high fit does not necessarily mean good forecasting ability.

In practice, it is important to consider some of these criteria as guide towards consistent and reasonable forecasts.

Interpretation of results

Several statistics are used to explain in a relatively simple terms the necessary steps in using econometric analysis to help in policymaking and to enhance the understanding of claim frequency variable in insurance ratemaking.

It is obvious from the results (see Tables below) that the models have all performed reasonably well given the simplicity of their functional forms. The economic variables all have positive slope coefficients and are statistically significant at the 5% level of significance. That is a general increase in payroll which is normally a function of economic activity will lead to an increase in expected frequency. The positive sign on the unemployment rate is as expected since it reflects the conjecture that workers tend to file more claims during hard times in the labor market. Kahley (2000) provides some reasons to support this in California. This is a question of moral hazard and can be significant where the unemployment benefits are relatively low compared to the workers compensation benefits. In general, the trend variable has a negative sign and it is statistically significant which means that there was a long-run downward tendency in frequency in California. This may be attributable to factors such as safety measures, technical progress, etc. The meaning and practical application of the coefficients on the trend variable will be discussed in more detail later in this paper.

Table 1: Results of Indemnity Claims Frequency. Note that employment is in 100 000 workers.

Models Coefficients for Indemnity			
	Linear	*Multiplicative*	*Exponential*
Intercept	-135.82(-2.67)	-66.7100(-4.32)	-0.2750(-.27)
Employment	1.6(3.54)	4.2510(4.54)	0.028(3.95)
Unemp Rate	3.16(4.39)	0.5398(6.07)	0.0726(5.27)
Trend	-1.06(-7.07)	-0.0259(-9.56)	-0.0246(-8.41)
DW	1.9567	1.9451	1.9385
Adj. R^2	0.66	0.7928	0.7491
AIC	193.04	-106.85	-104.61
SBC	201.36	-98.53	-96.30

Table 2: Results of Medical Claims Frequency

Models Coefficients for Medical			
	Linear	*Multiplicative*	*Exponential*
Intercept	-127.94(-3.57)	-46.35(-5.35)	1.4235(-.27)
Employment[4]	1.58(6.24)	3.0925(5.88)	0.22(3.95)
Unemp Rate	0.5943(1.40)	0.0892(6.07)	0.0129(5.27)
Trend	-1.3918(-14.89)	-0.020(-14.18)	-0.020(-14.40)
DW	1.9907	1.9835	1.9906
Adj. R2	0.9381	0.9316	0.9323
AIC	202.79	-113.91	-113.91
SBC	220.11	-105.60	-105.60

The models have a high within sample predictive or explanatory power. The coefficient[5] of determination is used to judge the explanatory power of the regressors. For the time period considered, the economic factors together with the time trend explained about 67-95% of the variation in

[4] Employment in 100 000 workers.

[5] The implicit R-squares are calculated for the nonlinear models to make them comparable since the dependent variables are not the same.

frequency. It is clear that the linear model has the lowest explanatory power while the log-linear and semi-log are almost indistinguishable. Also, the R-squares are generally higher for the medical frequency compared to the indemnity frequency. After correcting for first order serial correlation[6], the DW statistic improved significantly in all the models. They are all close to 2.0 which is an improvement from the barely 1.0 before correction. In general, it appears that serial correlation is a menace in claims or frequency data.

The interpretation of the regression coefficients is also a very important part of econometric modeling. From the table of results (Table 1 and 2), for the linear specification, a one unit change in employment (unit is 1000 employees) holding other factors constant will result in 0.02 unit change in both indemnity and medical frequency. Also, for unemployment rate, a unit (%) change will result in a 3.16 unit change in frequency. The trend variable is negative indicating a small but persistent downward development in frequency. It may mean that over time, claims tend to decline due to improvements in factors such as technology and the manufacturing/service mix of the labor market.

Since the specification of model 2 automatically yields percent changes, the interpretation of the constant elasticity model is that a one percentage change in employment will result in respectively 4% and 3% changes in indemnity and medical frequency. The trends are discussed later in the paper. The interpretation of the exponential model requires some special attention. Note that, a unit change in employment will result in 0.0002 % change in frequency for both the indemnity and medical frequency. It is easy to see that, the coefficient measures the relative change in frequency for a unit absolute change in the independent variables.

Model Selection Techniques

In practice, e.g., in actuarial trending procedures, one is often saddled with the question of which model is preferable to some other model(s). According to the ASP, the actuary should 'be familiar with and consider various methods in statistics and numerical analysis for measuring trends". This also entails steps for evaluating the tentatively selected model and possibly revising the model. This in fact means that the actuary is not opposed to new and improved methods of model selection.

---

[6] Serial correlation is when errors in one time period are correlated directly with errors in ensuing period.

There are many statistics that may help make an objective and consistent selection among competing models. The adjusted R-square has often been used to select models for forecasting purposes. This statistic has sometimes been criticized for not adequately imposing adequate penalty for the degrees of freedom. Thus some modern criteria such as the Akaike Information (AIC) and Schwartz Bayesian criteria (SBC) have been proposed. In employing these criteria for judging model's performance, the smaller the value of the statistic the better. We shall use different analytical model selection criteria in deciding which model is best for forecasting. From the tables (1 and 2) of results, it is again clear that the two nonlinear models are the winners, i.e. both the log-linear and the exponential models have smaller AIC and SBC than the linear model. The log-linear model has the smallest AIC and SBC. Note that, it is generally accepted that when AIC and SBC conflict, one should choose the model with lowest SBC.

Forecasting Frequency

At several levels of insurance business, decisions have to be made. To guide decision makers forecasts are often produced. For example, forecasts of expected claims (pure premium) are required in making rates. Under the credibility approach, the premium estimate for a loss if full credibility is applied is the average loss from the experience. Forecasts of trends have always been used as inputs in ratemaking process.

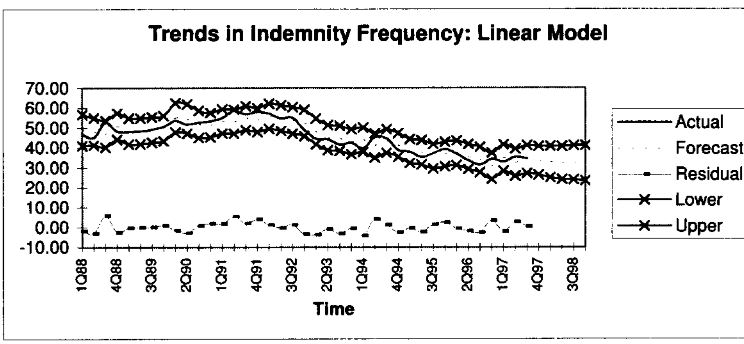
Understanding the steps in obtaining reasonable models is vital to improving the quality of the predictions and their application in real world. In light of these compelling reasons, we attempt to briefly explain the procedure using data from the California Workers Compensation Institute. To put our models to test, two types of forecasts are performed that is ex-post and ex-ante predictions. In the ex-post forecast, all values of the dependent and independent variables are known. This uses a subsample of the data to fit the model and then compares its forecast against the known remaining values. While in the ex-ante or conditional forecast all the variables are not known with certainty, forecasts of the input variables are used to produce the corresponding forecasts of the dependent variable.

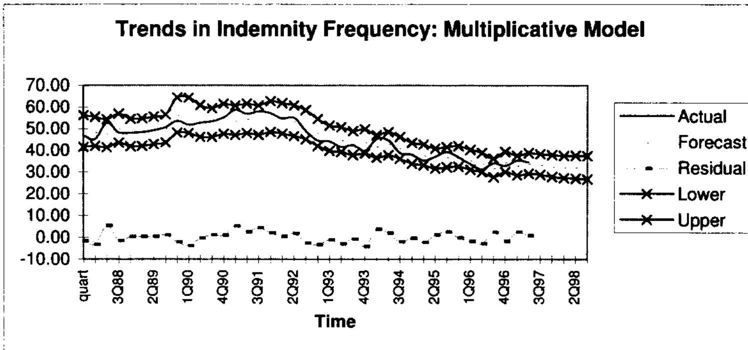
Note that we forecast the dependent variable(s) conditional on the assumptions of the independent variables. It is thus clear that any assumption about the input into the model affects the forecast generated.

As part of the rigorous process of model building and validation, we conducted an ex-post analysis of the models. The results are not discussed here but the general conclusion is that the nonlinear models have smaller root mean square errors of forecast. We discuss the results of the conditional forecasts in more detail. The plots below have the observed, the predicted, and the lower and upper confidence intervals of the frequency. The residuals are also provided which are found at the bottom of the graphs. Based on these plots, it seems quite apparent that, the nonlinear models have a much better fit and lower confidence bounds than the linear model. This is true in both the indemnity and medical cases.

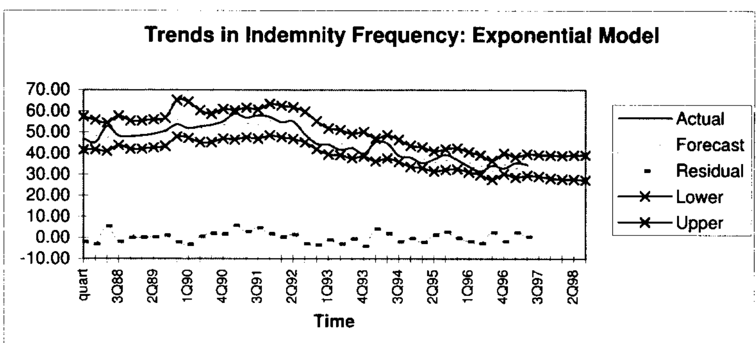
Also, the fits are much better in medical than in indemnity. Again, judging from the graphs, the nonlinear models are preferred to the linear specification. Note that the confidence bands are broader for the out of sample forecast reflecting the uncertainty in the model inputs. It appears that the linear model is much more sensitive to uncertainty in the input than the nonlinear ones. The interpretation of the confidence interval is that we are almost 95% confident that the realization of claims frequency for five quarters hence will fall within the confidence limits. As pointed out earlier, the real test of these models is when we get the data for the five quarters and compare them with the predictions for each of the models.

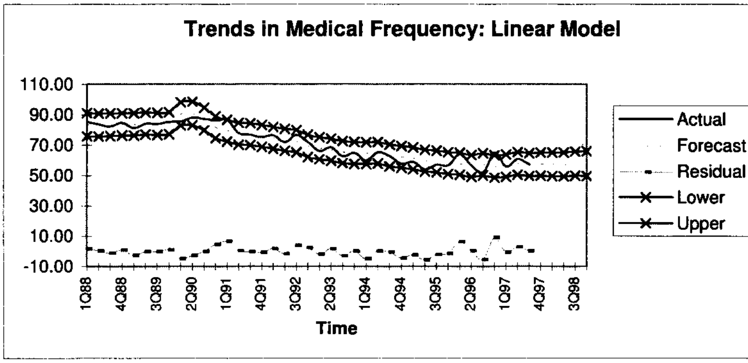
Graphs of actual frequency and predicted versus time in quarters.



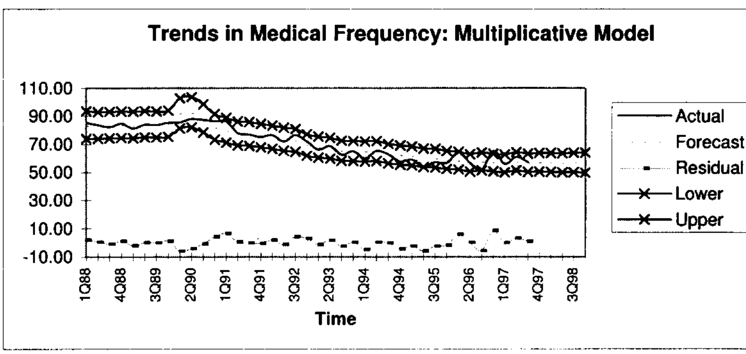


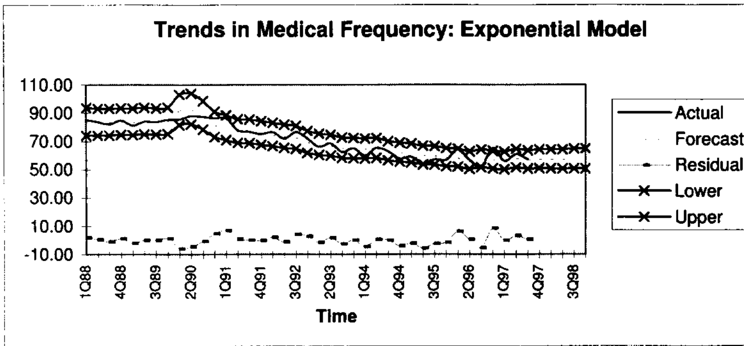
Graphs of actual frequency and predicted versus time in quarters.





Graphs of actual frequency and predicted versus time in quarters.





Notes on Trend Forecasting: - Actuarial versus Econometric Trends

This section discusses some issues related to frequency trends in insurance ratemaking process. Actuarial trending procedures employ what is known as deterministic trend models in estimating trends in frequency, severity and loss ratios. Deterministic trend models are often used in other areas such as economics, engineering and finance. Commonly used models in actuarial ratemaking are the linear and exponential time trend models. The econometric models developed are extensions to the trend models used by actuaries [see equations (1)-(3)]. In econometric parlance, the actuarial models are said to be nested in the econometric models. These equations reduce to the actuarial models when linear identifying restrictions, i.e., $\beta_2 = \beta_3 = 0$ are applied. This restriction is tested in all the models assuming that the null hypothesis is, H_0: the actuarial model is preferable to the econometric model. It is then possible to use the likelihood ratio statistic[7] which is approximately chi-square distributed

---

[7] The statistic is calculated as
$2(\log likelihood\ big \bmod el - \log likelihood\ small \bmod el)$.

with degrees of freedom equal to the number of restrictions imposed or the F-statistic to test the hypothesis. We employ both test statistics in this analysis. The tests have all been very highly significant leading to the rejection of the null hypothesis even at 1% level of significance. Hence, it can be concluded that using additional economic variables is worthwhile.

To answer some interesting practical questions, we use the data to compute some trend indications in two different ways. They will be labeled actuarial and econometric trends respectively. This is just for the purposes of taxonomy but not more. Note that there is no any good reason why, as far as we know, anyone of these procedures will be judged completely superior to the other. Much will depend on the intent and purpose of the analysis and who in fact conducts the analysis.

The results are reported for all three models and for both indemnity and medical claims frequency. Note that, actuarial trend models have been run and trend estimates are calculated to compare the actuarial trend forecasts and the econometric forecasts. The approach taken here is the actuarial methods. Notice that, while the two nonlinear models produce the percent growth rates directly, some actuarial or economic adjustments need to be made to the coefficient of the linear model in order to calculate the trend indication. The adjusted results will be reported for the linear models to ease comparison.

Tables 3 and 4 show the trend estimates using the Actuarial and Econometric models. Note that, the linear model estimates do not give the trend estimates prior to credibility. However, after some actuarial or economic adjustments, the slope coefficients can be converted into annual percentage changes. In economics, such adjustments include the calculation of the percent change at the mean or some other statistics. Here, an economic judgment is required by a trained and experienced individual, i.e. the judgment must be informed one. The growth rate at the mean is calculated as the trend coefficient of the linear model divided by the mean frequency over the entire series. The average values are respectively 45.5359 and 71.3026 for indemnity and medical frequency. Thus the growth rate is –0.0114[8] for the actuarial indemnity model. The results also include the standard errors of estimation which can be used to construct confidence intervals for the trend indication. In the case of the nonlinear models, the indications can be calculated by exponentiating the estimated coefficient of the trend variable. For example, using actuarial

[8] That is –0.5248 divided by 45.5359 equals linear trend. The other values can be calculated in a similar fashion.

estimate for indemnity will give an indication of 0.9878[9] while the econometric exponential model gives 0.9754. The econometric multiplicative model gives a slightly different number from the exponential, i.e. 0.9738. Similarly, this calculation can be done for the medical frequency models. Note that in estimating the coefficients, the number of observations is large compared to what actuaries would normally have available. Statistically, the more observations used, the smaller are the errors since they tend to cancel out. However, practical limitations and experience may warrant the use of new and recent observations.

[9] This is calculated as exp (-0.0122).

Table 3 Trend Estimates: Estimates from the linear and nonlinear actuarial models.

	Linear	*Multiplicative*	*Exponential*
Indemnity	-0.0114(0.0019)	--------	-0.0122(0.0018)
Medical	-0.0141 (0.001)	--------	-0.0142(0.00108)
IndRSquare	0.5115	--------	0.5489
MedRSquare	0.8438	--------	0.8364

Note that in Table 3, the multiplicative and the exponential estimates are the same in this case.

Table 4 Trend Estimates: Estimates from the linear and nonlinear econometric models.

	Linear	*Multiplicative*	*Exponential*
Indemnity	-0.0237(0.0035)	-0.0265(0.0028)	-0.0249(0.0031)
Medical	-0.0193 (0.0014)	-0.0204(0.0015)	-0.0203()
IndRSquare	0.6224	0.7682	0.7094
MedRsquare	0.9268	0.9243	0.9223

It is not surprising that the econometric models gave much better fit than the actuarial counterparts. In econometrics, it is expect that the bigger model will be at least as good as the smaller one but it is the magnitude of the gains that is dramatic in this case. As explained earlier these models use additional explanatory variables related to the economic demographic and social factors. The trends also show bigger declines in general than the actuarial model estimates. In actuarial trending procedures, several factors are taken into account. These include but not limited to goodness of fit measure, success of the model in making prior projection, etc. Thus a good model with a high explanatory power is welcome.

Summary and Conclusion

The study is conducted using the CWCI published quarterly data and three variants of econometric frequency models for both indemnity and medical experience. They all show that economic activity, i.e. the business cycle is still an important determinant of frequency. Most, if not all, previous studies are consistent with this observation. The important but not new message to practitioners and management is that economic activity is a significant determinant of claims frequency. Even in a declining frequency period, the decline may be slower than it would be in boom periods compared to periods of economic stagnation or recession.

Statistical tests have confirmed that nonlinearity is important in insurance claims frequency for the state of California Workers Compensation systems. This is because the two nonlinear models seem to outperform their linear counterpart. This makes sense because the real world itself is full of nonlinear relationships. In addition, it is shown that using economic variables resulted in a substantial payoff in terms of statistical performance. Trends have been calculated using actuarial and econometric models and the results have been discussed.

There are some theoretical issues regarding uncertainties in the input variables of an econometric model such as model uncertainty, variable and parameter uncertainty which have not been discussed in detail in this paper but are being considered as possible extensions to explore in future research.

Acknowledgments:

I thank Chun, Ruby, Tonya and Ahmed Ustarz for their help. I also thank Ziv for interesting and simulating discussions.

References

Kahley, W. J. California Workers' Compensation Claims Frequency Forecast, California Workers Compensation Institute, 2000.
Pindyck, R.S. and Rubinfeld, D.L., 1997, 4th Edition.
Ussif, A.M., An international analysis of workplace injuries, 2002.

A Discussion of "Risk Load for Insurers" by Sholom Feldblum

Trent R. Vaughn, FCAS, MAAA

DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXVII

RISK LOADS FOR INSURERS

BY SHOLOM FELDBLUM

DISCUSSION BY TRENT R. VAUGHN

Acknowledgment and Caveat Emptor

I would like to acknowledge the comments of the reviewers of this paper. In particular, the reviewers stated that the paper was "unnecessarily argumentative". And ... they may be right about that! I must confess that I do enjoy a good actuarial debate, and, in this case, I strongly believe that Feldblum's risk load methodology represents an unsound application of financial theory to the insurance problem. Even so, let me temper the criticisms below with a couple of caveats. First, an experienced actuary has pointed out to me that Feldblum was not alone in advocating this "modified CAPM" approach to risk loads. At the time that this paper was written, the approach was fairly common. Second, it is important to temper any criticisms of this paper with an acknowledgment of the many contributions that Feldblum has made to the actuarial and insurance literature. Certainly, Feldblum has contributed more to our profession than most actuaries (including myself) could achieve in several lifetimes.

When applying a financial theory to an insurance problem, one should logically follow several rules. First, one should carefully consider the proof of the financial theory, and determine whether the proof makes sense in the insurance setting. Second, one should consider the underlying message of the financial theory, and determine the implications of this message to the insurance problem. Finally, one should be aware of the empirical evidence in support (or contradiction) of the original financial theory. Feldblum's application of the CAPM to the insurance problem fails with regard to the first and second of these rules.

For instance, the actual proof of the CAPM relies on several key assumptions that are incomprehensible in describing the insurance company's choice between writing various lines of business. As an example, the original CAPM proof *assumes that the individual investor can supplement his purchases of marketable securities by borrowing or lending at the risk-free rate of interest, resulting in a linear efficient investment frontier. When an insurance company writes a policy, it invests the policy premium (and supporting surplus) in a variety of financial instruments, including risk-free bonds and/or risky common stocks. From this standpoint, it's unclear how an insurance company faces a choice between writing an insurance policy and borrowing or lending at the risk-free interest rate. As a result, the logic and proof underlying the CAPM cannot rationally be applied to the insurance company's portfolio problem.*

In addition, Feldblum's proposed formula is inconsistent with the very message of the original CAPM. The original CAPM is predicated on the fact that individual investor's can (and do) reduce their risk via individual portfolio diversification. Individual investor diversification is completely outside the scope of Feldblum's formula. Instead, Feldblum's formula measures the "risk" of a given insurance policy only in terms of the insurance company's underwriting portfolio. It does not consider the fact that individual investor's do not hold the insurance company's stock in isolation, but as part of a well-diversified investment portfolio.

In the spirit of full disclosure, let me point out that this paper was soundly rejected for publication in the Proceedings. *As such, I do appreciate the* CAS Forum *as a venue for unique ideas. But, as with all* CAS Forum *articles: let the buyer beware!*

1. INTRODUCTION

Feldblum's paper "Risk Loads for Insurers" discusses various methodologies for estimating the insurance risk load. According to this paper, traditional methods are inadequate. As such, the majority of the paper discusses a proposed methodology for applying modern portfolio theory and the capital asset pricing model (CAPM) to the insurance pricing problem.

Unfortunately, the proposed methodology represents an unsound application of financial theory to an insurance problem. Specifically, the proposed methodology merely borrows the notation of the CAPM, without considering the underlying assumptions and logic of the CAPM paradigm.

Section 2 of this paper will present an actual algebraic proof of the CAPM. In Section 3, we summarize the assumptions underlying the proof and discuss the implications of the result. Section 4 addresses Feldblum's methodology, and points out the unsound nature of that approach. Lastly, Section 5 describes a correct application of the CAPM paradigm to the insurance pricing problem.

2. ALGEBRAIC PROOF OF THE CAPM

Mossin [10] first provided an algebraic formulation of the CAPM proof. In this section, we will briefly outline the key elements of Mossin's proof.[1]

The Investor's Constrained Maximization Problem

The description of the investor's constrained maximization problem is primarily due to Markowitz [5] [6] and Tobin [12]. Markowitz first described the impact of portfolio diversification; Tobin extended the analysis by quantifying the investor's utility of wealth as a function of the mean and variance of total portfolio return.

Assume that there are m individual investors, $i = 1, 2, \ldots m$. Each of these investors possesses an initial wealth amount of w_i, which will be used to purchase securities. At the end of one-period, these securities will be sold, and the proceeds will be used to purchase goods and services for consumption. In other words, we are working with a one-period model of investor behavior. This is also sometimes referred to as a "two-date" model, since the investor purchases securities at time $t=0$ and sells these securities at time $t=1$.

Assume that there are n securities, $j = 1,2, \ldots, n$, each offering a total payment of D_j at the end of one period. Since we are considering a one-period model, D_j can be considered to be a liquidating dividend on the security. In addition, the total payment D_j will be distributed to the various security holders in proportion to the security holder's ownership stake in the firm. For instance, if an individual purchases 25% of the available amount of a security at time 0, then he will be entitled to 25% of the total liquidating dividend on that security at time 1.

Also, assume that for each $j = 1, 2, \ldots, n$, D_j is a normally-distributed random variable. The variance-covariance matrix, Σ, is assumed to be positive-definite; the properties of positive-definite matrices imply that there is no-risk free security (that is, $Var(D_j)>0$ for all $j = 1, 2, \ldots n$) and no two securities are perfectly negatively correlated (that is, the correlation coefficient for each pair of distinct securities is not equal to negative one). Moreover, in addition to the available

[1] Note: The proof in this section is not exactly identical to Mossin's original proof. We have modified the notation, changed the order, and added several clarifying remarks.

market securities, assume that each investor can borrow or lend at the risk-free rate of interest r_f.

At the beginning to the period, each investor must decide how to allocate his available wealth among the various securities. For instance, let x_{ij} be the proportion of the total issue of security j that is purchased by investor i.. In addition, we will let d_i represent the total dollar amount that the investor lends at the risk-free-rate.[2] At the end of the period, the total payment T_i received by investor i will be given by the following expression:

$$T_i = \sum_{j=1}^{n} x_{ij} D_j + d_i(1 + r_f).$$

Thus, for each investor, the total payment at the end of the period is a normally distributed random variable. The mean and variance of this random variable are given by the following expressions:

$$E(T_i) = \sum_{j=1}^{n} x_{ij} E(D_j) + d_i(1 + r_f)$$

$$Var(T_i) = \sum_{j=1}^{n} \sum_{k=1}^{n} x_{ij} x_{ik} Cov(D_j, D_k)$$

Let each investor's utility of end-of-period wealth be given by the function $u_i(w)$, i = 1,2, ... ,m. Moreover, we will assume that each investor is risk-averse (that is, $d^2u_i(w)/dw^2 < 0$) and maximizes the expected value of his utility of end-of-period wealth. Tobin [12] demonstrated that, under these assumptions, each investor's expected utility of end-of-period wealth is a function solely of the mean and variance of the investor's total end-of-period payment. That is, expected utility of end-of-period wealth is given by $E[u_i(T_i)] = f_i[E(T_i), Var(T_i)]$.

Hence, at the beginning of the period, each investor solves a constrained maximization problem. Specifically, each investor will maximize $f_i[E(T_i), Var(T_i)]$ subject to the following wealth constraint:

$$\sum_{j=1}^{n} x_{ij} v_j + d_i = w_i,$$

where v_j represents the total market value of security j.

---

[2] If the investor borrows at the risk-free rate in order to purchase additional securities, then $d_i < 0$.

Prior to solving this constrained maximization problem, several specific assumptions should be emphasized. First, we are assuming that the following inputs are all exogenous to the model: (1) the risk preferences of each of the m individuals, as given by their utility of end-of-period wealth functions, (2) the payoff characteristics of the n securities, and (3) the risk-free rate of interest r_f. Second, we are assuming that all assets are marketable and infinitely divisible. Third, we are ignoring taxes and transaction costs. Fourth, we are assuming that investors have homogenous expectations regarding security returns, and that each investor can borrow and lend as much as he wishes at the same risk-free rate of interest. Lastly, we are assuming perfectly competitive security markets; this assumption implies (among other things) that each investor can purchase as much of each security as he wishes at the prevailing market price.

In order to solve the constrained maximization problem, we will utilize the method of Lagrange multipliers:

$$L = f_i[E(T_i), Var(T_i)] + \lambda_i(w_i - \sum_{j=1}^{n} x_{ij}v_j - d_i)$$

Taking partial derivatives with respect to x_{ij} (j=1,2, ... ,n) and d_i and equating them to zero yields:

$$\partial L / \partial x_{ij} = [\partial f_i / \partial E(T_i)][\partial E(T_i) / \partial x_{ij}] + [\partial f_i / \partial Var(T_i)][\partial Var(T_i) / \partial x_{ij}] - \lambda_i v_j$$
$$= [\partial fi / \partial E(T_i)]E(D_j) + [\partial f_i / \partial Var(T_i)]\sum_{k=1}^{n} 2x_{ik}Cov(D_j, D_k) - \lambda_i v_j = 0$$

$$\partial L / \partial d_i = [\partial f_i / E(T_i)][\partial E(T_i) / \partial d_i] - \lambda_i$$
$$= [\partial f_i / \partial E(T_i)](1 + r_f) - \lambda_i = 0$$
$$\Rightarrow \lambda_i = (1 + r_f)[\partial f_i / \partial E(T_i)]$$

Substituting λ_i into the first set of equations and rearranging yields the following:

$$[\partial f_i / E(T_i)][E(D_j) - (1 + r_f)v_j] = -[\partial f_i / Var(T_i)]\sum_{k=1}^{n} 2x_{ik}Cov(D_j, D_k) \quad (2.1)$$
$$\forall j = 1,2,....,n$$

Thus, each investor (i = 1,2, ... ,m) solves the above system of n equations for the n unknown variables x_{ij}, j=1,2,.....,n.

The Market Clearing Mechanism

Sharpe [11], Lintner [4], and Mossin [10] extended the above analysis to a market equilibrium setting. As each investor solves the above set of equations, security prices (and the resulting total market values, v_j) will adjust to accomodate imbalances between supply and demand. Equilibrium is reached when each investor solves the above equations and the market "clears" for each asset (this market clearing condition will be made more precise later). In this paper, we will ignore the conditions under which equilibrium is attained. Instead, we will assume that equilibrium is reached and examine the properties of the resulting equilibrium.

Arbitrarily select a given investor i and two distinct securities a and b. Taking the ratio of equation (2.1) for these two assets yields the following:

$$\{[\partial f_i / E(T_i)][E(D_a) - (1+r_f)v_a]\} \;/\; \{[\partial f_i / E(T_i)][E(D_b) - (1+r_f)v_b]\}$$
$$= \{-[\partial f_i / Var(T_i)]\sum_{k=1}^{n} 2x_{ik}Cov(D_a, D_k)\} \;/\; \{-[\partial f_i / Var(T_i)]\sum_{k=1}^{n} 2x_{ik}Cov(D_b, D_k)\}$$

After cancelling factors and rearranging terms, we have the following equality:

$$\sum_{k=1}^{n} x_{ik}Cov(D_a, D_k)/[E(D_a) - (1+r_f)v_a]$$
$$= \sum_{k=1}^{n} x_{ik}Cov(D_b, D_k)/[E(D_b) - (1+r_f)v_b] \qquad (2.2)$$

In order for the market to clear, the excess supply for each security must be zero. That is, the sum of the weights for each asset must equal 1. In symbolic terms, the market clearing condition is as follows:

$$\sum_{i=1}^{m} x_{ij} = 1$$
$$\forall j = 1,2,....,n \qquad (2.3)$$

By summing both sides of (2.2) across all investors (i = 1,2,...,m) then applying (2.3) and rearranging terms gives us the following:

$$[E(D_a)-(1+r_f)v_a]/\sum_{k=1}^{n}Cov(D_a,D_k)$$
$$=[E(D_b)-(1+r_f)v_b]/\sum_{k=1}^{n}Cov(D_b,D_k)=\Theta \qquad (2.4)$$

Summing equation (2.4) over all assets yields the following:

$$\sum_{j=1}^{n}[E(D_a)-(1+r_f)v_a]/\sum_{j=1}^{n}\sum_{k=1}^{n}Cov(D_j,D_k)$$
$$=[E(D_M)-(1+r_f)v_M]/Var(D_M)=\Theta, \qquad (2.5)$$

where D_M is total payment on the market portfolio and v_M is the total value of the market portfolio. That is, $D_M = D_1 + D_2 + ... + D_n$, and $v_M = v_1 + v_2 + ... + v_n$.

Combining (2.4) and (2.5) yields the following:

$$[E(D_a)-(1+r_f)v_a]/\sum_{k=1}^{n}Cov(D_a,D_k)=[E(D_M)-(1+r_f)v_M]/Var(D_M) \qquad (2.6)$$

Using the notation developed above for D_M, note that $\sum_{k=1}^{n}Cov(D_a,D_k)$ can be rewritten as $Cov(D_a,D_M)$. By using this revised notation, equation (2.6) can be solved for the value of an asset under market equilibrium:

$$v_a=\{E(D_a)-[Cov(D_a,D_M)/Var(D_M)][E(D_M)-(1+r_f)v_M]\}/(1+r_f) \qquad (2.7)$$

Equation (2.7) can be converted into rates of return by using the following:

$$R_a = (D_a - v_a) / v_a \qquad (2.8)$$
$$R_M = (D_M - v_M) / v_M \qquad (2.9)$$

Substituting (2.8) and (2.9) into (2.7) yields (after some algebra):

$$E(R_a)=r_f+[Cov(R_a,R_M)/Var(R_M)][E(R_M)-r_f] \qquad (2.10)$$

Equation (2.10) is the traditional Sharpe/Lintner/Mossin CAPM.

3. THE CAPM AS A PARADIGM

As demonstrated in the previous section, the proof of the traditional Sharpe/Lintner/Mossin CAPM is predicated on the following key assumptions:

1. Individual investors are risk averse and maximize their expected utility of end-of-period wealth.
2. Investors have homogenous expectations regarding securities with a joint normal distribution of total payments.
3. Investors can borrow and lend as much as they want at the risk-free interest rate.
4. Security markets are perfectly competitive.
5. All assets are marketable and infinitely divisible.
6. There are no taxes, transaction costs or restrictions on short selling.

By rearranging formula (2.10), the CAPM predicts that the equilibrium expected return on an individual asset a will be given by the following formula:

$$E(R_a)-r_f=Cov(R_a,R_M)\{[E(R_M)-r_f]/Var(R_M)\},$$

where R_a is the the return on asset a, R_M is the return on the market portfolio, and r_f is the risk-free rate.

The difference between the expected return on asset a and the risk-free rate, also known as the "risk margin", is thus seen to be the product of two terms: the "risk" of asset a (as given by $Cov(R_a,R_M)$) and the "market price of risk". A common explanation of this definition of "risk" is that investors are only compensated (via the risk margin) for undiversifiable, or "systematic", risk. Investors are not compensated for diversifiable, or "unique", risk. In other words, investors are not concerned about the variance of the asset's return if held in isolation; instead, investors are concerned only with the covariance of that asset's return with the overall market return.

With respect to this interpretation, you sometimes hear the following objection: how can investors ignore the variance of an individual asset's return when that variance contributes to the "risk" of the asset? After all, every asset is included in the market portfolio. Thus, $Var(R_a)$ is actually one of the terms in $Cov(R_a,R_M)$, and thus contributes to the risk premium in asset a's expected return.

This objection, however, is a trifling issue. In real-world security markets, the number of securities n is extremely large. In order to see this, let's re-write the "risk" of the asset, or Cov(R_a, R_M), as the following sum of n terms:

$$Cov(R_a, R_M) = (v_a / v_M)Var(R_a) + \sum_{j \neq a}(v_j / v_M)Cov(R_a, R_j),$$

where, for each security j (j = 1,2, ... ,n), the ratio v_j/v_M is the relative value of asset j as a percentage of the value of the entire market portfolio. In this manner, Var(R_a) is only one of a very large number of terms and v_a/v_M is likely to be very close to zero. As a result, Fama and Miller [2] note that "the variance term in the asset's risk is likely to be trivial relative to the weighted sum of covariances."

Here we can draw an analogy to classic microeconomic price theory. Under the theory of perfect competition, we assume that the individual firm is a price taker; this firm faces a horizontal demand curve and can sell as many units as it wishes at the prevailing market price. If we consider the demand curve for market as a whole, however, price is inversely related to the quantity produced. But isn't each individual firm part of the overall market? How, then, can it sell any given quantity at a fixed price?

The solution to this conundrum lies in the specifications of the economic model; in the model of perfect competition, we require a very large number of producers (and buyers), with no one producer comprising a significant proportion of the overall market. In this case, the actions of any one producer will produce only a negligible impact on the overall market price. Likewise, if apply the CAPM model to a world with a very large number of securities, each security's variance has only a trifling impact on its risk.

4. APPLYING THE CAPM NOTATION TO INSURANCE PRICING

In the past, practicing actuaries have been tempted to borrow the results of the CAPM and apply this notation to insurance pricing. Meyers [7, p.4] describes the rationale as follows: "It would seem desirable to adapt this securities pricing model to the insurance pricing problem. One possible approach would be to let an insurer play the role of the investor and let an insurance policy, or a line of insurance, play the role of the individual security and use the CAPM directly."[3]

---

[3] As an aside, Meyers adds (in a footnote on the same page), "This is the approach taken by the so called 'Insurance CAPM', which is described in 'Asset Pricing Models for Insurance' by J. David Cummins, ASTIN Bulletin, November 1990, p. 125." It is important to note, however, that Cummins definitely does not use this approach

Feldblum uses this general approach in his paper. Feldblum's methodology essentially applies the CAPM notation to the insurance pricing problem. Feldblum summarizes his approach as follows:

"An insurer chooses lines of insurance (or blocks of business) to maximize its expected return while minimizing its 'risk'. The market return R_m in the CAPM model should be replaced by the return on a fully diversified insurance portfolio. The appropriate equation is $R = R_f + B(R_p - R_f)$, where R_p is the return on the all lines combined insurance portfolio."

In addition, Feldblum derives each line's "beta" by a regression between the operating returns on that line and the operating returns for all property/liability insurance lines combined. Thus, Feldblum's full formula coincides with the Sharpe/Lintner/Mossin CAPM formula, but with an "insurance interpretation" of the variables:

$$E(R_a) - r_f = Cov(R_a, R_P)\{[E(R_P) - r_f]/Var(R_P)\}$$

But is this formula really sound? In a recent PCAS paper, Mildenhall [9] describes the difference between applying a *paradigm* and simply borrowing a *notation*. The approach above simply borrows the CAPM notation while ignoring the major underlying message of the CAPM paradigm. As such, the technique clearly does not represent a logical extension of financial theory to insurance pricing.

Specifically, the CAPM is a paradigm that describes equilibrium in a capital market with risk-averse individuals and a large number of assets. As discussed in the previous section, the main result of the CAPM is that risk-averse investors are only concerned about the systematic, or undiversifiable, risk of individual assets. In this case, corporations, including insurance companies, will not be "risk-averse" in the same sense as individual investors. On the contrary, the CAPM implies that corporations are not concerned about the total variance of results, but only the extent to which these results fluctuate in step with overall economic conditions. Ironically, by simply applying the notation (or framework) of the CAPM to the insurance pricing problem, one is implicitly contradicting the very message of the CAPM paradigm.

The major cause of these problems is that the underlying logic and proof of the CAPM do not apply to the insurance company's choice between individual insurance policies or lines of business. As noted above, the CAPM requires strictly

---

anywhere in his paper. Instead, Cummins uses a correct application of the CAPM paradigm to an insurance pricing, which will be described in Section 5 of this paper.

risk averse individual investors. In addition, the fundamental CAPM result hinges on the assumption of a large number of assets, as demonstrated in the previous section. Meyers [7] points out a major flaw with replacing the thousands of individual securities in the original CAPM with only a few lines of business in the insurance problem; namely, with only a small number of lines of business, the variance of each line contributes significantly to that line's "beta" (and thus to its risk margin), contradicting a key implication of the CAPM that only undiversifiable risk is relevant.

This problem can be seen clearly by returning to our original analogy from microeconomics. The idea that each individual firm is a price taker hinges on the assumption of a large number of competing firms. Consequently, if we apply the model of perfect competition to a product market with only 15 firms, the underlying logic falls apart.

Furthermore, a closer look at the actual CAPM reveals other assumptions that may need to be modified before applying the proof to insurance markets . In particular, the CAPM assumes that the individual investor incurs no transaction costs in the process of forming a diversified portfolio. This assumption may be reasonable for individual investors, as mutual funds offer extensive diversification in exchange for a relatively low expense charge. Insurance companies, however, incur much more extensive transaction costs in the process of forming a diversified portfolio of insurance policies. Likewise, security markets are generally viewed as perfectly competitive, given the large number of both buyers and sellers, and the widespread availability of information. Insurance markets, on the other hand, may not always be perfectly competitive; for instance, there may only be a handful of insurance companies operating in certain "niche" lines. [4]

Lastly, it is unclear how certain assumptions in the actual CAPM proof even apply to the insurance market. As an example, the CAPM assumes that there are no restrictions on short selling; in symbolic terms, an investor "shorts" a security j by selecting an x_{ij} factor that is less than zero. But how does an insurance company "short" a given line of insurance? Also, the CAPM assumes that the investor has the option of supplementing his purchases and sales in marketable securities by borrowing or lending at the risk-free rate of interest. But what meaning does this have in relation to an insurance company's choice between writing various lines of insurance? When an insurance company writes a policy, it invests the premium in various financial instruments, including risky common stocks and risk-free

[4] The complications of transaction costs in insurance markets and a limited number of lines of business were part of the motivation behind the Competitive Market Equilibrium risk load formula, developed by Meyers [7]. Also, see Meyers [8] for a related discussion of the flaws in Feldblum's methodology.

government bonds. It is unclear how the insurance company faces a "choice" between writing insurance and borrowing or lending money.[5]

5. A CORRECT APPLICATION OF THE CAPM PARADIGM TO THE INSURANCE PRICING PROBLEM

A correct application of the CAPM paradigm to the insurance pricing problem reflects the underlying message of the CAPM: individual investors hold diversified portfolios and only require compensation (in the form of a higher expected return) for undiversifiable risk. In other words, we must recognize that individual investors do not hold the insurance company's common stock in isolation, but only as a small part of a well-diversified portfolio. Hence, the risk margin on the insurer's common stock return is proportional to the "beta" of that common stock, or the extent to which it varies with the overall return on the market portfolio.

The major implication of the CAPM to insurance pricing is that we can't consider the insurer's underwriting results in isolation, because individual investors hold insurance stocks as part of a well-diversified portfolio. Thus, the required return on an insurance company's common stock depends on the correlation between the stock's return and the return on the market portfolio.

As noted in a footnote above, Cummins [1] describes the correct application of the CAPM to the insurance pricing problem. In this formulation, one determines the "fair" premium, or the premium that equates the expected rate of return to the required rate of return. Moreover, the required rate of return is determined in accordance with the CAPM, by examining the correlation between the return on the insurance policy and the return on all securities in the financial marketplace.

Of course, the difficulty in correctly applying the CAPM to the insurance pricing problem involves the necessary parameter estimation. Fortunately, Garven [3] has shown that the option pricing method is consistent with the CAPM approach, while allowing for easier estimation of the necessary parameters.

---

[5] Actuaries do occasionally attempt to estimate insurance portfolios that lie on the "efficient frontier". These estimates do not typically include the line tangent to the risk-free rate and the efficient frontier, as is commonly done in the estimation of the efficient frontier of financial securities.

6. SUMMARY

As the financial and insurance sectors continue to consolidate, actuaries are becoming exposed to a myriad of financial theories. As we progress into the 21st century, we will be required to apply these financial theories to problems in insurance. As we complete this endeavor, it is critical to avoid making the same mistakes of the past. A common mistake, as demonstrated in Feldblum's paper, is to simply "borrow" the notation of a financial theory, without considering the assumptions, logic, and implications of the underlying paradigm.

Misapplications of this nature result in more than just bad theory; they also sow widespread confusion. For instance, after reading Feldblum's paper, actuaries will be tempted to partition the total risk of an insurance line of business into two components: the portion that is explained by the variation of operating returns on all insurance lines combined, and the portion that is due specifically to the unique attributes of the line under consideration. The first of these components may be labelled "systematic risk", and the second "unique risk". Using this terminology, systematic risk represents the risk that an insurance company cannot eliminate via diversification across various lines of business.

In the financial world, however, systematic risk represents the underlying risk of a security that an individual investor cannot eliminate via portfolio diversification. In the study and application of finance, the concept of systematic risk pertains to an individual investor's diversification across securities, not to a corporation's attempt to diversify across lines or divisions.

Moreover, the logical framework of the CAPM implies that investors will be rewarded (in a linear manner) only for the risk that cannot be eliminated by individual diversification. There is no sound basis for applying the CAPM to a corporation's choice between various lines (or divisions), and stating that the expected return on a given line will be linearly proportional to the risk that cannot be eliminated via corporate diversification. The two reasons for this are as follows: (1) the assumptions and proof of the CAPM do not apply to a corporation's choice between divisions or lines, and (2) the fundamental message of the CAPM is that individual investor's can diversify on their own; hence corporate diversification is redundant.

REFERENCES

[1] Cummins, J. David, "Asset Pricing Models for Insurance Ratemaking," *ASTIN Bulletin*, November 1990, pp. 125-166.

[2] Fama, Eugene F, and Miller, Merton M., *The Theory of Finance*, 1972.

[3] Garven, James R., "An Exposition of the Implications of Limited Liability and Asymmetric Taxes for Property-Liability Insurance," *Journal of Risk and Insurance*, September 1990, pp. 391-430.

[4] Lintner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, February 1965, pp. 13-37.

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

[6] Markowitz, Harry M., *Portfolio Diversification: Efficient Diversification of Investments*, 1959.

[7] Meyers, Glenn G., "An Introduction to the Competitive Market Equilibrium Risk Load Formula," pp. 1-28.

[8] Meyers, Glenn G., "Review of 'Risk Loads for Insurers' PCAS LXXVII (1990 by Sholom Feldblum)", *CAS Forum*, Winter 1996, pp. 85-96.

[9] Mildenhall, Stephen J. "Discussion of 'Application of the Option Market Paradigm to the Solution of Insurance Problems'," PCAS LXXXVII, 2000, pp. 162-187.

[10] Mossin, Jan, "Equilibrium in a Capital Asset Market," *Econometrica*, October 1966, pp. 768-783.

[11] Sharpe, William F., "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, September 1964, pp. 425-442.

[12] Tobin, J., "Liquidity Preference as a Behavior toward Risk," *Review of Economic Studies*, February 1958, pp.65-86.

Capital Allocation: An Opinionated Survey

Gary G. Venter, FCAS, MAAA

Capital Allocation: An OPINIONATED Survey

GARY G VENTER, GUY CARPENTER INSTRAT

A number of methods of allocating capital to business unit, e.g., line of business, profit center, etc., are discussed. Goals of capital allocation include testing the profitability of business units and determining which units could best be grown to add value to the firm. Methods of approaching these questions without allocating capital are included in the discussion.

CAPITAL ALLOCATION: AN OPINIONATED SURVEY

Capital allocation is generally not an end in itself, but rather an intermediate step in a decision-making process. Trying to determine which business units are most profitable relative to the risk they bear is a typical example. Pricing for risk is another.

Return-on-capital thinking would look to allocate capital to each business unit, then divide the units' profits by that capital. Of course if profit were negative, you would not need to divide by anything to know it is not sufficient. But this approach would hope to be able to distinguish the profitable-but-not-enough-so units from the real value-adders.

The same issue can be approached without allocating capital, using a theory of market risk pricing. The actual pricing achieved by each business unit can be compared to the risk price needed. This would require having a good theory of risk pricing, where the previous approach would depend on having a good theory of capital allocation. Since both are addressing the same decisions, both will be included in this survey. For those who like to phrase the issue as one of return on capital, the pricing method can be put into allocation terminology after the fact by allocating capital to equalize the ratio of target return to capital across business units.

Rating business units by adequacy of return is not necessarily the final purpose of the exercise. The rating could be used in further decisions, such as compensation and strategies for growth. For strategic decisions another question is important – not how much capital a business unit uses, but how much more is needed to support the target growth. In general it will be profitable to grow if the additional return exceeds the cost of the additional capital. In some cases a company might not need too much more than it already has for the target growth, in which case not much additional profit would be needed to make the growth worthwhile.

This is the marginal pricing approach, and is a basic tenet of financial analysis. It differs from capital allocation in that for marginal-cost pricing not all capital has to be allocated to reach a decision. Only the cost of the capital needed to support the strategy has to be determined, to see if it is less than the profit anticipated. Methods of quantifying the cost of marginal capital will be reviewed here as well, as again this is aiming at answering the same strategic questions.

Finally, another way to determine which business units are adding most to the profitability of the firm is to compare the insurer to a leveraged investment fund. Sometimes this is called the cost-of-float approach. The overall return of the insurer can be evaluated by finding the borrowing rate that would equalize its risk and return after tax to a leveraged investment fund. If the fund would have to be able borrow significant funds at a particularly low rate of interest to match the insurer's risk and return, then the insurance business is clearly adding value. The business units can be ranked based on their impacts on this borrowing rate.

Thus while the general topic is capital allocation, this survey is looking at methods for answering questions that capital allocation is addressing. To summarize, four basic approaches will be reviewed:

1. Selecting a risk measure and an allocation method and using them to allocate capital
2. Comparing actual vs. model pricing by business unit
3. Computing the cost of the marginal capital needed for or released by target strategies
4. Evaluating profitability in comparison to a leveraged mutual fund

The time period for evaluation is an issue for all of these methods, and this is addressed in Appendix 1.

APPROACH 1 – ALLOCATING VIA A RISK MEASURE

Table 1 lists a number of risk measures that could be used in capital allocation. To summarize briefly, VaR, or value at risk, is a selected percentile of the distribution of outcomes. For instance, the value at risk for a company might be the losses it would experience in the worst year in 10,000.

Table 1: Risk Measures

- ❑ **VaR**
- ❑ **EPD**
- ❑ **Tail VaR**
- ❑ **X TVaR**
- ❑ **Standard Deviation**
- ❑ **Variance**
- ❑ **Semi-Variance**
- ❑ **Cost of Default Option**
- ❑ **Mean of Transformed Loss**

EPD is expected policyholder deficit, i.e., the expected value of de fault amounts. It can also be generalized to include the expected deficit beyond some level, rather than beyond default. If b is the target amount, the EPD beyond b is: Pr(X>b)E[(X – b) | X>b].

Tail value at risk is the expected losses in the event that losses exceed the value-at-risk target. If the target loss level is b, this is E(X | X>b).

X TVaR is similar to Tail VaR, but rather than the mean of all cases over a level, it is the average for those cases of the excess of the losses over the overall mean, i.e., E[X– m | X>b].

A company with limited liability does not pay once its capital is exhausted. So the insurer holds an option to put the default costs to the policyholders. The value of this option can be used as a risk measure. The other measures are standard statistical quantities.

Often when allocating capital with a risk measure, the total capital is expressed as the risk measure for the entire company. For instance, the probability level can be found so that the Tail VaR for the company at that probability level is the capital carried. The capital could also be expressed as a multiple of the risk measure. For instance, the company could have a goal that the average loss

■ Table 2: Allocation Methods
❑ Proportional Spread
❑ Marginal Analysis
◆ By business unit
◆ Incremental by business unit
❑ Game Theory
❑ Equalize Relative Risk
❑ Apply Co-Measure

in the 1-in-100 year or worse not use up more than premium plus $^{1}/_{3}$ of capital. This would make the capital goal three times the 99% X TVaR. This is consistent with the idea that renewal business has a value, so the goal should be to have enough capital to continue operating even in the identified adverse situation. Also, some amount of capital might be set aside as not being risk capital – it could be for acquisitions perhaps – and the remainder used to calibrate the risk measure. In any case, once the total capital has been associated with a risk measure, an allocation method can be applied to get that capital split to the business unit level by allocating the risk measurement. Several possible allocation methods are given in Table 2. Not all of these work with all of the risk measures.

Proportional spread is the most direct method – apply the risk measure to each business unit and then allocate the total capital by the ratio of business unit risk measure to the sum of all the units' risk measures. Usually the sum of the individual risks will be greater than the total risk, so this method is crediting each unit with a diversification benefit.

Marginal analysis measures the risk of the company with and without a specified business unit. The difference in required total capital is then the marginal capital for the business unit. The total capital can then be allocated by the ratio of the business unit marginal capital to the sum of the marginal capital of all the units. This usually allocates more than the marginal capital to each unit. The incremental marginal method is similar, but the change in capital is calculated for just the last increment of expected loss for the unit, say the last dollar. Whatever reduction that is produced in the risk measure by eliminating one dollar of expected loss from the business unit is expressed as a capital reduction ratio (capital saved per dollar of expected loss) and applied to the entire unit to get its implied incremental marginal capital to use in the allocation.

The game theory approach is another variant of the marginal approach, but the business units are allowed to form coalitions with each other. The marginal capital for a unit is calculated for every group of units it could be a part of, and these are averaged. This gets around one objection to marginal allocation – that it treats every unit as the last one in. This method is sometimes called the Shapley method after a founder of game theory.

The Myers-Read method also uses marginal allocation. It sets the marginal capital needed to support an exposure increase equal to the additional capital it would take to make the cost of the default put, as a percentage of expected losses, the same before and after. It has the advantage over other marginal methods that the marginal increments add up to the total capital. This method is discussed in detail in Appendix 2.

Equalizing relative risk involves allocating capital so that each unit, when viewed as a separate company, has the same risk relative to expected losses. Applying this to the EPD measures, for instance, would allocate enough capital to each business unit make the EPD for every unit the same percentage of expected loss.

Co-measures were introduced by Rodney Kreps as a way of allocating capital in an additive manner that is nonetheless consistent with the overall risk measure used to define total capital. Appendix 3 discusses these in greater detail. They can be most easily thought of in terms of a scenario generator. Take the case where the total capital requirement is set to be the tail value at risk at the 1-in-1000 probability level. Then in generating scenarios, about 1 in 1000 would be above that level. The co-Tail VaR for each business unit would just be the average of its losses in those scenarios. This is its contribution to the overall Tail VaR.

Co-measures provide a totally additive allocation. Business units could be combined or subdivided in any way and the co-Tail VaR's would add up. For instance, all the lines of business could be allocated capital by co-Tail VaR, then each of these allocated down to state level, and those added up to get the state-by-state capital levels for all lines combined. This could be done for peril or other business categories as well.

Commentary on Allocation by Risk Measures

VaR could be considered to be a shareholder viewpoint, as once capital is exhausted, the amount by which it has been exhausted is of no concern to them. EPD, default option cost, X TVaR, and Tail VaR relate more to the policyholder viewpoint, as they are sensitive to the degree of default. And indeed the shareholders might do well when they consider policyholder needs. All of these measures ignore risk below the critical probability selected. VaR also ignores risk above that level, while the tail measures evaluate that risk linearly, which many consider to be an underweighting.

Variance does not distinguish between upward and downward deviations, and so could provide a distorted view of risk when these directions are not symmetric – which is the usual case. Semi-variance looks only at adverse deviations, so accounts for this. Taking the mean of a transformed loss distribution is a risk measure aiming at quantifying the financial equivalent of a risky position, and it can get around the problems of the tail methods. More exploration of transformations could be useful.

Allocating by marginal methods is accepted in financial theory. However, allocating more than the pure marginal capital to a unit it could lead to pricing by a mixture of fixed and marginal capital costs, violating the marginal pricing principle. Even when the total

capital is the sum of the marginal increments, as in Myers-Read, there is no tie-in between the capital allocated to a line and the value of its risk. Thus it would be a great coincidence if this allocated capital were right for a return-on-capital ranking.
The co-measure approach is consistent with the total risk measure and is completely additive. Thus if the risk measure gives the right capital need overall, the co-measure shows each line's contribution to that. But it too could violate marginal pricing.

Myers-Read was introduced as a method of allocating the frictional costs of holding capital. These are discussed more in Appendix 2, but as a definition I would propose that costs which arise from holding capital even if no risk is written are frictional costs. Corporate tax on investment income is an example. A more delicate issue is any lower investment income resulting from taking less investment risk in order to give policyholders greater security. I would hold that this is a frictional cost as well. Even though it results from the intent to sell insurance, this does not differentiate it from other frictional costs.

The return for actually putting the capital at risk is a different matter. This relates to the amount of risk taken, not the amount of capital allocated. In financial models beta is almost always a component of the return for bearing risk, but it is not generally a part of the frictional cost. Some actuarial pricing approaches have assumed that pricing to recoup frictional costs is sufficient, and this is encouraged by assertions that beta is zero for underwriting anyway. More recent theory, discussed below, shows that risk pricing is more than beta. This suggests that even if allocating capital by risk measure is sufficient for allocating frictional costs, there are other elements of return that will not be proportional to the amount of capital held and so should be measured in some other way.

APPROACH 2 – COMPARE ACTUAL VS. MODEL PRICING

A traditional use of capital allocation is to price business to equalize return on capital. However even if the allocation method is intuitively satisfying, there is no guarantee that such pricing would correspond to the market value of the risk transfer. If instead actual pricing were compared to value pricing, the profitability of business units could be evaluated without allocating capital at all (except to the degree this is necessary in the pricing to compute the frictional costs of holding capital). But for those who still prefer a single target return on capital, capital could be allocated after the pricing by equalizing the return on capital from the value prices.

This method requires an evaluation of the market value of the risk transfer provided. Financial methods for valuing risk transfer typically use transformations of the loss probabilities to risk-adjusted probabilities, with covariance loadings like CAPM being one special case. This is a fairly technical calculation and to date there is no universal agreement on how to do it. Some transforms do appear to give fairly good approximations to actual market prices, however. The Wang transform has been used successfully in several markets to approximate risk pricing. Finance professionals now appear to favor an adjusted CAPM approach that corrects many of the over-simplifications of the original formulation. For instance, a correlation with the insurer's own results may be as important as correlation with the market in determining the cost of risk transfer.

To use CAPM or similar methods, costs are first identified, then a risk adjustment added. Three elements of cost have been identified for this process: loss costs, expense costs, and the frictional costs of holding capital. The latter is not the same as the reward for bearing risk, which is separately incorporated in the risk adjustment.

The CAS Committee on the Theory of Risk is sponsoring the Risk Premium Project to look into how to do risk pricing right. Starting from CAPM, they are looking at are several considerations needed to get a realistic market value of risk transfer. Some issues in this area are:

- Company-specific risk needs to be incorporated, both for differential costs of retaining vs. raising capital[1] and for meeting customer security requirements.
- The estimation of beta itself is not an easy matter[2]
- Other factors besides beta are needed to account for actual risk pricing[3]
- To account for the heavy tail of P&C losses, some method is needed to go beyond variance and covariance[4,5]
- Jump risk needs to be considered. Sudden jumps seem to be more expensive risks than continuous variability, possibly because they are more difficult to hedge by replication. Large jumps are an element of insurance risk, so need to be recognized in the pricing.

Commentary on Target Pricing

Measures of the market value of risk transfer are improving, and even though there is no universally accepted unique method, comparing actual profits to market-risk-model profits can be a useful evaluation. This can then be reformulated as a capital allo-

---

[1] Froot, Kenneth A. and Stein, Jeremy C., *A New Approach to Capital Budgeting for Financial Institutions*, Journal of Applied Corporate Finance, Summer 1998, Volume 11, Number 2

[2] Kaplan, Paul D. and Peterson, James D., *Full-Information Industry Betas* Financial Management 27 2 Summer 1998

[3] Fama, Eugene F. and French, Kenneth R. *Multifactor Explanations of Asset Pricing Anomalies* Journal of Finance 51 1 March

[4] Wang, Shaun *A Universal Framework For Pricing Financial And Insurance Risks*, ASTIN Bulletin, 2002, Volume 32, No. 2

[5] Kozik, Thomas J. and Larson, Aaron M. *The N-Moment Insurance CAPM*, Proceedings of the Casualty Actuarial Society LXXXVIII, 2001

cation if so desired. The pricing can also be particularized to the company, considering that company costs of risk transfer may differ from the industry's. However the requisite pricing models are still under development.

APPROACH 3 – CALCULATING MARGINAL CAPITAL COSTS

A third approach to evaluating business unit profitability is to look at the last increment of business written by the unit to see whether the cost of the additional capital required is less than the profit it generates. This is not necessarily an allocation of capital, in that the sum of the marginal increments may not add up to the total capital cost of the firm, leaving some fixed capital not allocated. It does correspond, however, to the financial principle of marginal pricing. In basic terms, if the profit from adding an increment of business in a unit exceeds its marginal capital cost, then the unit should be expanded.

Because of the unallocated fixed capital charges, an anomalous situation could arise where each business unit is profitable enough on the margin but the firm is not so as a whole. In such cases further strategic analysis would be needed to reach an overall satisfactory position for the firm. One possibility might be to grow all the business units enough to cover the fixed charges. Another might be to look a merger possibilities.

One way to do the marginal calculation would be to set a risk requirement for overall capital, and then see how much incremental capital is needed to continue to meet this requirement after the small expansion of the unit. This is the same approach used in the incremental marginal capital allocation by risk measure, but there is no allocation. The cost of capital would be applied to the incremental capital and compared directly to the incremental expected profits.

Another way to calculate marginal capital costs is the options-based method introduced by Merton and Perold. A business unit of an insurer could be regarded as a separate business operating without capital, but with a financial guarantee provided by the parent company. If the premium and investment income generated by the unit is not enough to pay the losses, the firm guarantees payment, up to its full capital. In return, if there are any profits, the firm gets them.

Both the value of the financial guarantee and the value of the profits can be estimated using option pricing techniques. The financial guarantee in effect gives the unit's policyholders an option that allows them to put any losses above the unit's premium and investment income to the firm. But this is not unlimited, due to the firm's limited resources, so the value of this guarantee is the difference between two put options: the option with a strike at losses equal to the sum of premium plus investment income, less the value of the insolvency put. The firm's call on the profits is a call option with strike of zero. If that is worth more than the financial guarantee provided, the business unit is adding value. These options would take some work to evaluate, however, in that the lognormal assumption of Black-Scholes would often be not sufficiently heavy-tailed. The options pricing could also reflect the specific cost to the firm of providing the guarantee, which would take into account guarantees provided to correlated business units. The managers of the unit could also be treated as having a contingent claim on the profits through incentive compensation.

Commentary on Marginal Capital Costs

This method directly evaluates marginal costs of decisions, so it can correctly assess their financial impact. If a large jump in business – upwards or downwards – is contemplated, the marginal impact of that entire package should be evaluated instead of the incremental marginals. There is still a potential arbitrary step of the criteria chosen for the aggregate capital standard, however. This is avoided in the financial guarantee approach, but that is more difficult to calculate, in that some method of pricing heavy-tailed options would be required.

APPROACH 4 – MUTUAL FUND COMPARISON

An insurer can be viewed as a tax-disadvantaged leveraged mutual investment fund. It is tax-disadvantaged since a mutual fund does not usually have to pay tax on its earnings. It is leveraged in that it usually has more assets to invest than just its capital. An equivalent mutual fund can be defined as one that has the same capital and the same after-tax probability distribution of returns as the insurer. It can be specified by its borrowing rate, the amount borrowed, and the investment portfolio. This should provide enough variables to be able to find such a mutual fund. If there are more than one such, they could all be considered as strategic alternatives and the easiest one to create would be the equivalent.

The insurer can be evaluated by the equivalent borrowing rate. If the investors can duplicate the risk and return by not writing insurance but by borrowing at a high rate of interest, there is not much value in writing the insurance, as they could readily borrow the money instead. However if they have to be able to borrow at a very low rate to get an equivalent return, the insurer is producing a result that is not so easily replicated by a leveraged mutual fund.

This is first of all a method for evaluating the overall value added of the insurer, but it can be done excluding or adding a business unit or part of a business unit to see if doing so improves the comparison. If a business unit lowers the equivalent borrowing rate on the margin, making a loan more difficult to get by the equivalent mutual fund, it is increasing the value of the firm.

Commentary on Mutual Fund Comparison

This is a potentially useful analysis, but it would require modeling the distribution function of return for the entire firm, including all risk and return elements, and a potentially extensive search procedure for finding the equivalent mutual fund.

CONCLUSIONS

The allocation method in the end depends on why you are allocating capital. Allocating by a risk measure is straightforward but subjective. It appears to be appropriate for allocating frictional capital costs, which are proportional to capital, but not for return on risk bearing, which might not be. If it also allocates fixed costs, it could produce misleading indications of actual profitability prospects. Strong candidates for risk-measure allocations are Myers-Read and co-X TVaR. Both start with reasonable stories of the overall capital need – enough to keep the default cost low for MR and enough to be able to continue writing after the very bad year for X TVaR. Then they both allocate all the capital in an additive manner which directly reflects the individual contributions to the overall capital need. The capital standard for MR sounds a little stronger in theory, but the computational aspects are harder than they might appear. The value of the put involves calculations way out in the tail of a distribution whose tail is not known that precisely. X TVaR can use a capital standard for partial loss of surplus, which is more reliably modeled than default.

Pricing comparison is applicable to evaluating the actual realized pricing including frictional and risk transfer costs. However, it is only as good as the pricing model used, and that could be complicated.

The marginal cost method shows directly the impact of growing each business unit. It still requires a choice for the overall capital standard, unless the financial guarantee method is used, in which case it requires an appropriate option pricing formula.

The mutual fund comparison could be computationally intensive, but would provide qualitative insight into the value of the firm and its business units.

References

Kreps, Rodney E, *Riskiness Leverage Models*, Instrat working paper, to appear.

Merton, Robert and Perold, Andre "Theory of Risk Capital in Financial Firms," *Journal of Applied Corporate Finance*, Fall 1993.

Myers, Stewart C and Read, James A. 2001, "Capital Allocation for Insurance Companies," *Journal of Risk and Insurance*, 68:4, 545-580.

Appendix 1: Time Frame for Evaluation

Different business units will tend to pay their losses out over different time frames. This complicates the capital cost allocation issues. Generally speaking, capital will be needed to support reserves as they run off, and this should get into the allocation. More research would be useful to specify how to do this in each approach. An outline of some possibilities for this is below.

It is possible to quantify the remaining runoff risk for each year for each business unit. The years would be correlated, as issues in the claims environment could hit several years at once. Methods using risk measures could incorporate this runoff risk. To put the years together, a cost of capital could be applied to each year, and then discounted. Ongoing investment income on premiums not yet paid out could be discounted as well. This could be done historically on existing reserves or prospectively on the projected payout pattern.

Pricing transformations could use a similar approach. The adjusted probabilities for the cash flow stream could be transformed and discounted. One way Myers-Read could adapt to this is by considering a sequence of default put options – one at each year end as policies run off. These become increasingly more likely to be hit as the time frame expands. The prices of these options could be present-valued and summed up to get a total value of the default puts for current writings. Then for a small increase in writings in any business unit, the additional capital needed to keep this total put value constant, as a percent of expected losses, could be calculated and used as the basis of capital allocated to the unit. The marginal amounts seem likely to add up to total capital, as they would for each of the annual puts separately.

A similar method should work for pricing in the financial guarantee approach. The firm could be getting a sequence of call options and providing a sequence of put options, whose total prices could be compared.

For the mutual fund comparison it would seem sufficient to look at the current annual risk to earnings including runoff risk for current liabilities. This would not be a totally prospective look at current strategies, but would still provide a valuable perspective on the financial status of the firm as it has been managed to date.

APPENDIX 2: THE MYERS-READ APPROACH

Myers-Read capital allocation presents a challenge to the classification of methods, in that it allocates all capital, it provides a marginal capital cost, and it can be used in pricing. But in the context of ranking returns, it is a risk-measure based method.

Butsic provides a slightly different derivation of the allocation formula than do Myers-Read themselves, and his approach is basically followed here, referred to as MR. You can get the same result from slightly different sets of assumptions, so this is not one of those situations where if you accept the assumptions you must accept the result. The results and assumptions can be evaluated from various viewpoints, and so the question is, does the whole approach work well?

The context for the method is that there are frictional costs to holding capital. In some countries, insurer investment income is subject to taxation, so tax is a frictional cost in those jurisdictions. Unless the insurer has really vast amounts of capital, it often has to invest more conservatively than the owners themselves would want to, due to the interests of policyholders, regulators, and rating agencies. There is a liquidity penalty as investors cannot get their investments out directly, and there are agency costs associated with holding large pools of capital, i.e., an additional cost corresponding to the reluctance of investors to let someone else control their funds, especially if that agent can pay itself from the results.

MR assumes a pricing approach in which the policyholders are charged for these frictional costs. This requires that the costs be allocated to the policyholders in some fashion, and MR uses capital allocation to do that. Every policyholder gets charged the same percentage of its allocated capital for the frictional costs. Thus it is really the frictional costs that are being allocated, and capital allocation is a way to represent that cost allocation. The formula can be adapted to include in the premium other risk charges that are not proportional to capital, so this capital allocation does not necessarily provide a basis for a return-on-capital calculation.

A key element of the MR approach is the value of the default put option. As a company with limited liability, an insurer does not pay losses once its capital is exhausted. So it can be said that the insurer holds an option to put the default costs to the policyholders. MR assumes a lognormal or normal distribution for the insurer's entire loss portfolio, so can use the Black-Scholes options pricing formula to compute D, the value of this put option.

Adding a little bit of exposure in any policy or business unit has the potential to slightly increase the value of the default option. But adding a little more capital can bring the value of this option back to its original value, when expressed as a percentage of total expected losses. The MR method essentially allocates this additional capital to the additional exposure that required it.

In other words, the default option value, as a percentage of expected losses, i.e., D/L, is held as a fixed target, and the last dollar of each policy is charged with the amount of extra capital needed to maintain that target option value. But any dollar could be considered the last, so the whole policy is charged at the per dollar cost of the last dollar of expected loss. The beauty of the method is that those marginal capital allocations add up to the entire capital of the firm.

In the MR development, the total capital requirement of the firm is never specified, but could be taken to be the amount of capital needed to get D/L to a target value. The allocation method is the incremental marginal effect method – the incremental dollar loss for the business unit or policy is charged with the amount of capital needed to keep D/L at its target.

The total capital is the sum of the individual capital charges, i.e., $\sum c_iL_i = cL$, where c_iL_i is the capital for the ith policy with expected losses L_i , and cL is total capital. Thus each policy's (or business unit's) capital is proportional to its expected losses, and the capital allocation question becomes how to determine the allocation factors c_i.

Formally, MR requires that the derivative of D with respect to L_i be equal to the target ratio D/L for every policy. Butsic shows that this condition follows from some standard capital market pricing assumptions. This requirement means that the marginal change in the default cost due to a dollar (i.e., fixed, small) change in any policy's expected losses is D/L. Thus D/L does not change with an incremental change in the expected losses of any policy. How is this possible? Because increasing L_i by a dollar increases capital by c_i, which is set to be enough to keep D/L constant when L_i increases. Thus the formal requirement that $\partial D/\partial L_i = D/L$ means that the change in c_iL_i due to a small change in L_i has to be enough to keep D/L constant.

The question then is, can allocation factors c_i be found to satisfy both $\sum c_iL_i = cL$ and $\partial D/\partial L_i = D/L$? That is, can by-policy capital-to-expected-loss ratios be found so that any marginal increase in any policy's expected losses keeps D/L constant, while the marginal capital charges sum to the overall capital? The MR derivation says yes.

In the MR setup, after expenses and frictional costs, assets are just expected losses plus capital, and so the Black-Scholes formula gives:

$$D = L[N(y+v) - (1+c)N(y)]$$

where v is the volatility of company results, $y = -\ln(1+c)/v - v/2$ and N(y) denotes the cumulative standard normal distribution.

Using this to expand the condition that $\partial D/\partial L_i = D/L$ requires the calculation of the partial of c w.r.t. L_i. Plugging in $\sum c_i L_i = cL$, this partial derivative turns out to be $(c_i - c)/L$. This leads to an expression for c_i in terms of c and some other things, which is the basis of the allocation of capital. This is how the condition on $\partial D/\partial L_i$ leads to an expression for c_i.

To express the allocation formula, denote the CV of losses as k_L and the CV of losses for the ith policy or business unit by k_i. Also define the policy beta as $b_i = \rho_{iL} k_i/k_L$, where ρ_{iL} is the correlation coefficient between policy i and total losses. Myers-Read also considers correlation of assets and losses, but Butsic gives the following simplified version of the capital allocation formula, assuming that the loss-asset correlation is zero:

$$c_i = c + (b_i - 1)Z, \text{ where } Z = (1+c)n(y)k_L^2/[N(y)v(1+ k_L^2)]$$

Butsic provides a simple example of this calculation. A company with three lines is assumed, with expect losses, CV's, and correlations as shown below. The total capital and its volatility are also givens. The rest of the table is calculated from those assumptions.

Changing the by-line expected losses in this example allows you to verify that if you add a dollar of expected losses to any of the lines, the overall D/L ratio is maintained by adding an amount to capital equal to the c_i ratio for that line.

Some aspects of the approach can be illuminated by varying some of the input assumptions. The examples that follow keep the volatility of assets constant, even though assets vary, which seems reasonable.

	line 1	line 2	line 3	total	volatilities
EL	500	400	100	1000	
CV	0.2	0.3	0.5	0.2119	0.2096
corr 1	1	0.75	0		
corr 2	0.75	1	0		
corr 3	0	0	1		
variance	10,000	14,400	2,500	44,900	
beta	0.8463	1.3029	0.5568		
capital	197.872	282.20	19.93	500	0.2209
assets				1500	0.0699
c_i:	0.3957	0.7055	0.1993	0.5	
- y:	1.9457807	**y+v:**	-1.7249		
N(y):	0.0258405	**N(y+v):**	0.042277		
n(y):	0.0600865	**1/n(y):**	16.64267		
Z:	0.6784		**D/L:**	0.0035159	

First, consider what happens if the CV for line 3 is set to zero. In this case, the line becomes a supplier of capital, not a user, in that it cannot collect more than it's mean, but it can get less, in the event of default. Then the capital charge c_i for this line becomes –17%, and the negative sign appears appropriate, given that the only risk is on the downside. The size of the coefficient seems surprising, however, in that its default cost is only 0.3% (which is the same for the other lines as well), but it gets a 17% credit. Part of what is happening is that adding independent exposures to a company will increase the default cost, but will decrease the D/L ratio, as the company becomes more stable. Thus in this case, increasing line 3's expected losses by a dollar decreases the capital needed to maintain the company's overall D/L ratio by 17 cents. This is the incremental marginal impact, but if line 3 decides to go net entirely, leaving only lines 1 and 2, the company will actually need $19.50 in additional capital to keep the same default loss ratio. This is the entire marginal impact of the line, which will vary from the incremental marginal.

Another illustrative case is setting line 3's CV to 0.335. In this case, its needed capital is zero. Adding a dollar more of expected loss maintains the overall D/L ratio with no additional capital. The additional stability from its independent exposures exactly offsets its variability. Again the marginal impact is less than the overall: eliminating the line in this case would require $10.60 in additional capital for the other lines.

The risk measure of the cost of the default option per dollar of expected loss, and the allocation principle that each dollar of expected loss be charged the frictional costs of the capital needed to maintain the target ratio, both appear reasonable, and the marginal costs adding up to the total eliminates the problem that fixed costs are being allocated using marginal costs. However, this is only so for incremental marginal costs. The marginal impacts of adding or eliminating large chunks of business can have a different effect than the incremental marginals, and so such proposals should be evaluated based on their total impacts.

Butsic also considers adding a risk load beyond the capital charge to the pricing. The same derivation flows through, just with expected losses replaced by loaded expected losses, and the capital charge set to c_i times the loaded losses. This provides a pricing formula that incorporates both risk load and frictional capital charges.

Using this, business unit results can be evaluated by comparing the actual pricing to the target pricing. If management wants to express this as a return on capital, the MR capital would not be appropriate. Rather the total capital should be re-allocated so that the ratio of modeled target profit to allocated capital is the same for each unit. Then comparing returns on capital would give the same evaluation as comparing profits to target profits. MR capital allocation would be the basis of allocating frictional capital costs, but not for calculating return on capital.

Appendix 3: Co-Measures

Co-measures can be defined for any risk measure that can be expressed as a conditional expectation, which is most of them. Suppose a risk measure for risk X with mean m can be defined as:

$R(X) = E[(X - am)g(x) \mid \text{condition}]$ for some value a and function g.

Suppose further that X is the sum of n portfolios X_i each with mean m_i. Then the co-measure for X_i is:

$$\text{co-}R(X_i) = E[(X_i - am_i)g(x) \mid \text{condition}]$$

Here the condition is the same as in the definition of R, so it is a condition on X, not X_i. Since expectations are additive, the sum of the co-R's of the n X_i's is R(X).

Variance

As an example, take a=1 and g(X) =X – m, with any condition that is always fulfilled, like 0X=0. Then R(X) is the variance of X. Thus,

$\text{co-}R(X_i) = E[(X_i - m_i)(X - m)]$, which is the covariance of X_i with X.

Value at Risk

Value at risk at probability level q can be defined as:

$$E(X \mid F(X)=q)$$

This is just the qth quantile of the distribution. Then the co-VaR is:

$$E(X_i \mid F(X)=q)$$

This would be the average value of portfolio i when total losses are at the qth quantile.

Tail Value at Risk

For probability level q, take a=0 and g(x) = 1, with condition F(X)>q. If q=99.9%, R is TVaR at the 1-in-1000 level. Then:

$$\text{co-TVaR}(X_j) = E[(X_i \mid F(X)>q)]$$

This is the mean loss for the ith unit in the case where total losses are over the q^{th} quantile.

Expected Policyholder Deficit

As another example, consider the expected policyholder deficit, or EPD. If X is all years' losses unpaid, b is total assets, and S(b)=1 – F(b), then:

$$\text{EPD} = E[(X - b)S(b) \mid X>b]$$

This is the R(X) form with a = 1, g(x) = S(b)(X – b)/(X – m) and condition X>b. With these, the co-measure is:

$$\text{Co-EPD}(X_j) = E[(X_i - m_i)g(X) \mid X>b]$$

$$= E[S(b)(X - b)(X_i - m_i)/(X - m) \mid X>b]$$

Each gets a fraction of the overall expected deficit given by the ratio of its losses above mean to the total losses above mean when there is a deficit.

Excess Tail Value at Risk

Define the measure excess tail value at risk by:

$$\text{X TVaR}_q = E[X - m \mid F(X)>q] \text{, so}$$
$$\text{Co-X TVaR}_q = E[X_i - m_i \mid F(X)>q]$$

If capital is set by X TVaR, it would provide enough to cover losses above mean losses for the average of the years in which losses exceeded the qth quantile. The capital allocated by Co-X TVaR to a line would be the line's average losses above its mean losses in those same adverse years. There should be some probability level q for which X TVaR or a multiple of it makes sense as a capital standard, as the mean loss should be already collected in premium. Using co-X TVaR for allocation would not charge capital to a unit for its mean losses. If by some chance the unit did not have losses above its mean in the average of the scenarios above the q^{th} quantile for the entire company, it would not be charged any capital. This makes sense if capital is indeed being held for the adverse outcomes.