The Casualty Actuarial Society *Forum*

Spring 2003 Edition

Including the Reinsurance Discussion Papers

To CAS Members:

This is the Spring 2003 Edition of the Casualty Actuarial Society *Forum*. It contains seven Reinsurance Discussion Papers.

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

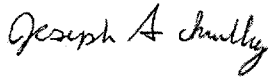
The CAS *Forum* is edited by the CAS Committee for the Casualty Actuarial Society *Forum*. Members of the committee invite all interested persons to submit papers on topics of interest to the actuarial community. Articles need not be written by a member of the CAS, but the paper's content must be relevant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

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

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

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

Sincerely,



Joseph A. Smalley, CAS *Forum* Chairperson

The 2003 CAS Reinsurance Discussion Papers
Presented at the
2003 Reinsurance Seminar
June 2-3, 2003
Sheraton Society Hill
Philadelphia, Pennsylvania

The Spring 2003 Edition of the CAS *Forum* is a cooperative effort between the CAS *Forum* Committee and the CAS Committee on Reinsurance Research.

The CAS Committee on Reinsurance Research presents for discussion seven papers prepared in response to its Call for 2003 Reinsurance Discussion Papers.

This *Forum* includes papers that will be discussed by the authors at the 2003 CAS Reinsurance Seminar, June 2-3, 2003, in Philadelphia, Pennsylvania.

Reinsurance Discussion Papers

Data Management Call Paper*

* First published in the Winter 2003 *Forum*; republished in its entirety.

The Valuation of Stochastic Cash Flows

Leigh J. Halliwell, FCAS, MAAA

The Valuation of Stochastic Cash Flows

Leigh J. Halliwell

Abstract

A stochastic cash flow depends on a random outcome. Since insurance and reinsurance contracts in exchange for financial considerations provide financial compensations against random outcomes, their compensations are perfect examples of stochastic cash flows. This paper develops a theory for the valuation of such cash flows from the four principles of present value, utility, optimum, and equilibrium. The most important implications of the theory are that the optimal amount for an economic agent to purchase depends on price, that (price, amount) loci are preliminary to market value, and that market value is the unique price at which all interested economic agents purchase optimal amounts. The theory belies the prevalent practices of risk-adjusted discounting and capital allocation (Appendix A), faulting them for naïve and erroneous conceptions of time. Accordingly, although the information that the theory presumes of economic agents is formidable, each agent is realistically burdened with ascertaining its own present opinions and preferences, rather than impossibly burdened with omniscience. And it sets the agents to the virtuous task of extracting value from projects, rather than from one another. The theory lays claim to fundamental principles of financial economics; it derives from the work of European risk theorists Karl Borch, Hans Bühlmann, and Hans Gerber, and gains support from a small but growing number of American actuaries. Though the paper remains theoretical throughout (especially in Appendices B-D), it furnishes several examples of sufficient detail for actuaries to apply it to pricing traditional insurance and reinsurance contracts.

Keywords: present value, utility, optimum, equilibrium

The Valuation of Stochastic Cash Flows

Leigh J. Halliwell

1. INTRODUCTION

The insurance and reinsurance industry is in an impasse. The muddled and contradictory thinking on how to value its contracts is the stuff of a Dilbert cartoon. Industry leaders act inconsistently, belittling theory as stifling action while grasping at every idea of the month. Actuaries too are swept along with the tide. What is needed, and what this paper hopes to accomplish, is to define basic economico-financial concepts and to develop a theory for the valuation of insurance and reinsurance contracts.

In the next section we define the basic concepts of stochastic cash flow and present value within the framework of modern probability theory. Then, in Section 3, we argue that present value is a random variable with a probability distribution, something that strikes at the heart of current financial theory. From there we turn in Section 4 to utility theory, for it alone seems to provide a criterion for preferring one probability distribution to another. But Section 5 gives a new twist to utility theory, according to which one purchases at a given price the amount that maximizes one's expected utility, as opposed to determining the price for a unit amount that maintains that utility. So the issue of price remains open until in Section 6 we see how it is collectively determined. Sections 5 and 6 contain what might aptly be called a market-tempered utility theory. Section 7

illustrates this theory with a reinsurance example, and suggests how to dampen the swings of the reinsurance market. Before the conclusion, Section 8 draws out two implications of the theory, especially how it could change actuaries *per se* from followers to leaders.

The theory to be developed is complicated. However, it is no more complicated than the various risk modeling and capital allocation projects in which companies have for years invested large sums of money and time, so far with little success. Moreover, very little herein is new; as radical as it may seem, the theory is just culled from the best actuarial, financial, and economic sources, from sources of unquestionable orthodoxy.

2. DEFINITIONS AND COMMENTS

We begin with the unexciting but important task of defining four essential terms: *cash flow*, *present value*, *stochastic [cash flow]*, and *random variable*.

A *cash flow* is a mapping, or function, c from $[0,\infty)$ into the real numbers that describes how much cash has been received up to and including $c(t)$. It should have at most a countably infinite number of discontinuities. Sufficient for our purposes are discrete cash flows, which jump at times t_i and are flat elsewhere. Thus we can imagine a cash flow as a countable set of increments, and denote it as $c = \{(t_i, x_i): i = 1,2,\mathrm{K}\ \}$, where an ordered pair (t,x) represents the receipt of cash x (dollars) at time t (years) from now. By definition, the present is time zero, and the time coordinates must be nonnegative. In other words, cash flows have no memory of the past.

Present value is a function v from $[0,\infty)$ onto the half-open interval $(0,1]$. It states that the receipt of one dollar t years from now is equivalent to the receipt of $v(t)$ dollars now. Four properties of present value are that $v(0) = 1$, that v strictly decreases, that v is continuous, and that $v(t)$ approaches zero as t approaches infinity. Just as cash flows have no memory of the past, present value looks only forward, not backward.[1]

Theoretically, each economic agent could chose its own present-value function; but practically, prices at which U.S. Treasury securities are traded determine a market-accepted present-value function at any moment of time. For centuries investors have been accustomed to think of present value in terms of yield. However, since yields are not constant by maturity, investors should abandon them in favor of the discount function with its equivalence between one dollar at time t and $v(t)$ now.[2]

The present value of discrete cash flow c is $PV[c] = \sum_i x_i \cdot v(t_i)$, if it exists.[3] It is an operator that assigns a real number to a function; and it implies that cash flow c is equivalent to the receipt of $PV[c]$ now. It is a linear operator in that for cash flows c and d and constants α and β, $PV[(\alpha c + \beta d)(t)] = \alpha PV[c(t)] + \beta PV[d(t)]$. Henceforth we

[1] Inasmuch as accounting looks backward over an accounting period, one must take care not to foist accounting concepts onto present value. The currently popular "results monitoring" projects have not recognized this.

[2] Halliwell [2001, Appendix A] challenges the notion that money somehow works at the yield. If a five-year STRIP (zero-coupon Treasury security) sells at 85.10 for a yield of 3.29 percent per year and a ten-year STRIP sells at 63.38 for a yield of 4.64 percent per year, is it helpful to assert that ten-year money is working harder than five-year?

[3] For some countably infinite cash flows, the sum does not converge. But the present value of a finite cash flow must exist.

assume that at any given moment every economic agent has a present-value operator (whether its own or the market-accepted), and can value any cash flow. If in the next moment the agent's operator changes, it must recalculate the present values of its cash flows; however, it does not follow that the agent should attempt to adjust its present-value operator for the risk that the operator itself will change.[4]

The adjective '*stochastic*' means 'probabilistic' or 'pertaining to a probability system'. In the 1920s and 30s the Russian mathematician Andrei Nikolaevich Kolmogorov devised the modern theory of probability with its fundamental concept of a probability system. In the words of Pfeiffer [1978, p. 29]:

> A *probability system* (or probability space) consists of the triple:
>
> 1. A *basic space* S of elementary outcomes (elements) ξ
> 2. A *class* $\mathcal{E}$ *of events* (a sigma field of subsets of S)
> 3. A *probability measure* $P(\cdot)$ defined for each event A in the class $\mathcal{E}$ and having the following properties:
>
> (P1) $P(S) = 1$ (probability of the sure event is unity)
> (P2) $P(A) \geq 0$ (probability of an event is nonnegative)
> (P3) If $\mathcal{A} = \{A_i;\ i \in J\}$ is a countable partition of A (i.e., a mutually exclusive class whose union is A), then
>
> $$P(A) = \sum_{i \in J} P(A_i) \qquad \text{(additivity property)}$$

A simplified version of a discrete probability system will serve our needs. The basic space S will consist of a countable number of elementary outcomes ξ_i, for $i = 1, 2, \ldots$. The class $\mathcal{E}$ of events will consist of all the subsets of S, i.e., it will be the power set of S. But we will concern ourselves with the "elementary" events $\{\xi_i\}$, and henceforth will

[4] A change in one's present-value operator can have profound effects, good as well as bad. On the bad side, it may render one insolvent accountingwise. However, that the operator might change in the next moment has no effect on the value of a cash flow at the present moment. The ultimate inference from this is that present value and solvency are unrelated (Section 8).

ignore the distinction between elementary events and elementary outcomes, referring to both indiscriminately as ξ_i. The class of these mutually exclusive ξ_i is a countable partition of S, and the probability measure is determined by nonnegative $P(\xi_i)$ that sum to unity.

Each agent has its own probability system, or in our simplified version, its own basic space S and probability measure P. Many elementary outcomes will appear only in one agent's basic space. Even to compare elements of different spaces may be difficult. But should two agents share an elementary outcome, they are free to assign different probabilities to it. And one agent may know for certain something that another regards as uncertain. Probability systems are as subjective as human beings; it is difficult, as well as unnecessary, for anyone to pontificate whether one system is more accurate or more in tune with reality than another. One can only hope that reality is reasonable enough to reward agents according to how accurately they perceive it.

Two characteristics of probability systems need to be appreciated. First, a probability system is as momentary as a present-value operator. An economic agent is free to change its system at any moment. But for now that system defines all that is, and to require it to consider how it might change would be endlessly circular. And second, a probability system is timeless. For example, if a coin toss were an elementary outcome, it would not matter if the toss had already happened and the agent were just ignorant of the outcome, or if the toss were yet to happen.[5]

[5] So we do not need to resolve the paradox of Schrödinger's cat: Before one opens the box is the cat either 100% dead or 100% alive, or is it suspended in a probability distribution between life and death?

A *stochastic cash flow* C is a mapping from a basic space into the collection of cash flows. In the simplified version, such a mapping assigns to each ξ_i a cash flow c_i. One can transfer probability from the elementary outcomes to the cash flows. For any cash flow c, let $A = \{\xi \in S : C(\xi) = c\}$. Then $P(c) = P(A)$. If no outcome maps to c, then $A = \varnothing$ and $P(c) = 0$. Or if many discrete ξ_i map to c, but the probability of each is zero, then still $P(c) = 0$. In these cases c is an impossible cash flow, according to the probability system under consideration. At the other extreme, if every discrete cash flow with positive probability maps to c, then $P(c) = 1$, and the stochastic cash flow is degenerate. There is as little difference between degenerate stochastic cash flow C and cash flow c that it equals with certainty as there is between degenerate random variable X and the number x that it equals with certainty.

Strictly speaking, a *random variable* is a function from S into the real numbers, which function is also "a measurable function with respect to the class of events $\mathcal{E}$." [Pfeiffer, 1978, p. 376] Our simplified version permits us to ignore the subtleties of measurability; for us a random variable is merely a function from the countable basic space into the real numbers. And the probability that the random variable equals some number is the sum of the probabilities of the elementary outcomes that map to that number.

3. PRESENT VALUE AS A RANDOM VARIABLE

So, at length, we are ready to define the present value of a stochastic cash flow. If C is

a mapping from S into cash flows (i.e., a stochastic cash flow), and PV is a mapping from cash flows into the real numbers, then the present value of stochastic cash flow C is the composite mapping $PV \circ C$. Being a mapping from S into the real numbers, it is a random variable.

As a very simple example of a stochastic cash flow, let S have two elementary outcomes: $S = \{H, T\}$. This basic space would model the outcome of a coin toss. The event class $\mathcal{E}$ would equal $\{\varnothing, \{H\}, \{T\}, S\}$. A fair coin would have the probability measure $P(\{H\}) = P(\{T\}) = 0.5$, with $P(\varnothing) = 0$ and $P(S) = 1$. This triple $(S, \mathcal{E}, P)$ satisfies the definition of a probability system. Let two cash flows be the receipts of 120 dollars and 80 dollars one year from now; they can be expressed incrementally as $\{(1,120)\}$ and $\{(1,80)\}$. Let stochastic cash flow C map the heads outcome to the first flow, and the tails outcome to the second. According to the set-theoretic definition of a function or mapping, $C = \{(H, \{(1,120)\}), (T, \{(1,80)\})\}$. At the end of August 2002, a one-year Treasury STRIP traded at 98.71 dollars (on a par value of 100 dollars). Taking that as time zero, we derive the market-accepted $v(1)$ as 0.9871. Hence, $PV(\{(1,120)\}) = 120 \times 0.9871 = 118.45$ dollars, and similarly, $PV(\{(1,80)\})$ equals 78.97 dollars. And the present value of the stochastic cash flow is the function or mapping $X = PV \circ C$, where $X(H) = PV(C(H)) = PV(\{(1,120)\}) = 118.45$, and similarly with the tails outcome for a value of 78.97. X satisfies the definition of a random variable, and we are allowed to make the four probability statements:

$$P(X = 118.45) = P(\{H\}) = 0.5$$
$$P(X = 78.97) = P(\{T\}) = 0.5$$
$$P(X = 118.45 \text{ or } X = 78.97) = P(S) = 1$$
$$P(X \neq 118.45 \text{ and } X \neq 78.97) = P(\varnothing) = 0$$

The cumulative distribution function of X is:

$$F_X(x) = \begin{cases} 0 & x < 78.97 \\ 0.5 & 78.97 \leq x < 118.45 \\ 1 & 118.45 \leq x \end{cases}$$

Exhibit 1 presents a slightly less simple stochastic cash flow. Ten elementary outcomes determine ten cash flows. Each cash flow consists of one ordered pair (t,x), as in the previous example, and these pairs are graphed on a Cartesian half-plane with present-value isobars. Treasury STRIP prices on 30 August 2002 at various maturities provided the raw data to which a $v(t)$ curve was fitted. So the isobar that stems from $(0, x_0)$ plots the curve $x_0/v(t)$; in other words, the isobars are proportional to one another, and fill the half-plane. The present values of the cash flows fall within the [75,100] interval. Provided that the probability of each elementary outcome is ten percent, Exhibit 2 graphs the cumulative density function of the present value of the stochastic cash flow. We have not yet treated valuation, but already one might suspect that the value of this stochastic cash flow must be greater than seventy-five and less than one hundred.

The third and final example of this section deals with a corporate bond. On 01 October 2001 investors paid just over 747 million dollars for unsecured bonds issued by Tyson Foods. The face amount of the bonds was 750 million dollars, and the coupon was 7.25 percent per year, payable semiannually. Moody's Investors Service rated the bonds as

Baa3, the lowest of the investment-grade ratings. Exhibit 3 contains this information, along with Moody's default probabilities and recovery parameters.[6] For example, Moody's estimates that Baa3 bonds have a 1.28% chance of default within three years of issue. The chance of the Tyson Foods bond default over its five-year term is 2.79%. And when a company defaults, holders of its bonds usually recover some of the face amount. Moody's estimates that holders of defaulted senior unsecured bonds recover 44.62% ± 26.32% of the face amount.

The bottom-left part of Exhibit 3 shows the cash flow that the bondholders hope to receive, the full payment of interest and principal. According to Moody's default probabilities their hopes have a 97.21% chance of fulfillment. On the bottom-right side of the exhibit are present-value factors, $v(t)$, based on current STRIP prices. The non-defaulted cash flow has a present value of 618,046,875 (principal) and 247,614,404 (interest) for a total of 865,661,279 dollars. The bottom-middle part of the exhibit shows the value of the cash flow that defaults after five interest payments and returns at maturity forty percent of the principal. The present value of this cash flow is 247,218,750 (principal) and 130,372,559 (interest) for a total of 377,591,309 dollars. The exhibit is part of a spreadsheet that simulates default probabilities and recovery rates. The default probability is a uniform (0,1) random variable, such as the 1.00% in the exhibit, which is compared with the Baa3 default probabilities. And the recovery factor (e.g., 40.00%) is modeled as a beta random variable whose parameters α and β are matched to the mean and standard deviation of Moody's recovery rates. The

[6] Moody's revises these probabilities and parameters annually, and one can download them from the website riskcalc.moodysrms.com/us/research/defrate.asp.

spreadsheet simulated 10,000 cash flows and calculated their present values. Exhibit 4 graphs[7] the distribution of the present value, along with reference lines for the mean ($E[X]$ = 854,166,433) and the purchase price ($q[X]$ = 747,187,500). Therefore, the expected profit is 106,978,933 dollars. The present value is at its maximum of 865,661,279 dollars past the ninety-seventh percentile, and thereafter declines in a fairly straight line effectively to zero. The probability that the bondholders will lose money, i.e., $P(PV(X) < 747{,}187{,}500)$ is 2.73%. Nearly all the defaults constitute a loss of money.

These three examples illustrate the concept of the present value of a stochastic cash flow as a random variable. It differs from the common concept of present value as a number. The common concept would discount a stochastic cash flow at a rate of return that is greater than the risk-free rate. Typical investors, considering the Tyson Foods bond at the time of issue, would know that the present value of its cash flow, if it does not default, is 865,661,279 dollars. And they would know that discounting the non-defaulted cash flow at 3.87 percent per year results in this dollar amount. Now if the bond does not default, its stochastic cash flow is degenerate and equivalent to a simple cash flow; hence the correctness of the equation $PV(X|\text{no default})$ = 865,661,279 dollars. This, indeed, is the mode of the present-value distribution (97.21% probable according to Moody's statistics). But does it do justice to the remaining 2.79% for one to decide somehow to discount the non-defaulted cash flow at 7.48%, and offer the resulting amount of 747,187,500 dollars? If not, then Donald Mango [2003, footnote 8] is correct to write that "the method of risk-adjusted discounting ... represents an

[7] Lee [1988] introduced actuaries to this form of graphing a probability distribution, wherein quantiles of the distribution are plotted against its percentiles, a form now commonly called a Lee diagram.

example of 'overloading an operator,' piling additional functional burden onto what should be a single purpose operator."

We believe that risk-adjusted discounting, with its concept of present value as a number, short-circuits the stochasticity of present value, is a legacy from pre-probabilistic, deterministic ages, and may be responsible for many of the woes in the past two decades of the insurance and reinsurance industry.[8] The conclusion of this section is that deriving the present value of a stochastic cash flow as a random variable is necessary for valuing the cash flow. More pointedly, if the random variables of the present values of two stochastic cash flows are equal, then the values of the two flows must also be equal.

4. THE UTILITY OF A DISTRIBUTION OF PRESENT VALUE

The value of stochastic cash flow C depends on the random variable X that maps from elementary outcomes ξ_i of basic space S to the present values of its outcome-dependent cash flows $PV[c_i]$. The cumulative distribution function of X, viz., $F_X(x)$, is $P(X \le x) = \sum_{PV[c_i] \le x} P(\xi_i)$. The cumulative distribution function of a random variable maps from the real numbers into the unit interval [0,1]. Though closely associated with its random variable, it lacks information about the basic space. Hence, from this distribution alone, one would not know how X relates to other random variables.

[8] Some think that the deterministic concept approximates the probability concept, perhaps as Newtonian physics approximates special relativity. However, the deficiency is much more glaring, rather like that of Aristotelian physics in comparison with Newtonian. Appendix A critiques risk-adjusted discounting.

However, an economic agent could evaluate the sum of its stochastic cash flows, i.e., its total (present-valued) stochastic wealth W, from the distribution of W. Such an agent would have to express preferences; for example, whether a normally distributed wealth of mean 600 and standard deviation 100 is preferable to one of mean 550 and standard deviation 50. So an economic agent values not stochastic cash flows themselves, but rather the effect of such cash flows on its total stochastic wealth. Let W represent an agent's current stochastic wealth. Let X be a stochastic cash flow offered at price q. If the agent purchases X, its resulting wealth will be $W+X-q$. The agent needs a decision-making operator U that assigns real numbers to distributions of stochastic wealth, such that W_1 is preferable to W_2 if and only if $U(W_1) > U(W_2)$.

Karl Borch [1961, p. 248] briefly recounts the history of utility theory, particularly how it fell out of favor with late nineteenth-century economists, who deemed it too complex to be workable. Around 1900 the indifference theory of Vilfredo Pareto, which seemed to circumvent utility functions, found favor with many. But in the 1940s Von Neumann and Morgenstern [1972] resurrected utility theory by proving that every decision-making operator U that possesses certain reasonable properties implies a utility function.[9] Borch himself offers a proof [1961, pp. 249-251] that every operator U that satisfies two axioms implies a utility function. The first axiom, which is hardly disputable, is that U constitutes "a complete preference ordering over the set of all probability distributions." The second axiom is that if U deems two distributions as equivalent, it will also deem equivalent their linear combinations with a third distribution. He discusses criticism of

[9] See also Debreu [1987, §4.6] for a proof of the existence of an n-dimensional utility function for an economic agent that can form preferences about baskets of n goods.

the second axiom, but concludes, "... this general criticism does not concern ... insurance where the only events considered are payment of different amounts of money." [1961, p. 254]

So we will proceed on the assumption that an economic agent estimates both the present value of its stochastic wealth W and the present value X of the stochastic cash flow. Moreover, the agent has a utility function u, so that its decision whether to purchase the cash flow at price q is determined by the relation $E[u(W+X-q)] \sim E[u(W)]$. To those who argue that estimating random variables and utility functions is too much to ask of economic agents, there is a two-part rejoinder. First, economic decision-making is not an easy problem. But rather than to cut the problem down to the size of the economic agents, we ought to build the agents up to the size of the problem. And second, the capital-allocation alternatives have grown so complicated that one rightly wonders whether it be any less difficult to apply them than to apply utility theory.

The properties of a realistic utility function are well known.[10] The utility of wealth w, $u(w)$, should be strictly increasing, twice differentiable, and concave downward. These properties imply increasing utility and diminishing marginal utility, viz., that $u'(w) > 0$ and $u''(w) < 0$. To such functions applies Jensen's inequality, $E[u(W)] \le u(E[W])$, with equality if and only if W is a degenerate random variable. Most treatments of utility

[10] Some actuarial references favorable to utility theory are Bowers [1986], Bühlmann [1980], Gerber [1979 and 1998], Halliwell [1999 and 2001], Longley-Cook [1998], Mango [2003], Panjer [1998], Schnapp [2001], Sundt [1991], and Van Slyke [1995 and 1999]. Economic works, e.g., Debreu [1987], Dixit [1990], Duffie [1990], and Von Neumann [1972], inevitably begin with a utility-theoretic foundation.

theory consider only these three properties; however, it is reasonable to add that $\lim_{w\to\infty} u'(w)=0$, from which it also follows that $\lim_{w\to\infty} u''(w)=0$. Furthermore, since stochastic cash flows can be much larger than an agents' stochastic wealth, one cannot set lower and upper bounds on $W+X-q$. Hence, $u(w)$ should exist for all real w. Candidates for utility functions are either quadratic, power, logarithmic, or exponential. But of these four types, only the exponential function is appropriate for all real numbers. For the power and logarithmic functions are not defined for zero and negative numbers, and the quadratic function decreases after its vertex.

The decision whether to purchase stochastic cash flow X at price q hinges on the comparison $E[u(W+X-q)] \sim E[u(W)]$. For a small enough q (perhaps negative), the left side of the equation is greater, and the purchase is desirable. But as q increases, the left side decreases and there should be a unique q^* at which both sides are equal. At a price greater than q^* the purchase is undesirable. So it is true of stochastic cash flows in general what Borch writes about insurance losses:

> In insurance a basic assumption is that there will always exist a unique amount of money which is the lowest premium at which a company will undertake to pay a claim with a known probability distribution. This assumption establishes an equivalence between certain and uncertain events. [1961, p. 249]

Borch writes from the insurance standpoint of receiving a minimum positive premium p for paying positive losses L, from which standpoint the decision hinges on the comparison $E[u(W-L+p)] \sim E[u(W)]$. But paying a positive loss corresponds to receiving a negative payoff, and receiving a positive premium to paying a negative price. So "the lowest premium at which the company will undertake to pay a claim" from Borch's insurance standpoint is equivalent to the price q^* "greater than which the

purchase is undesirable" from our investment standpoint. It augurs well for a valuation theory that one may treat liabilities as negative assets, and assets as negative liabilities.

For one whose wealth is the real number w, the decision to purchase hinges on the comparison $E[u(w+X-q)] \sim u(w)$. And due to Jensen's inequality:

$$\begin{aligned} E[u(w+X-q)] &\leq u(E[w+X-q]) \\ &\leq u(w+E[X]-q) \end{aligned}$$

If q equals $E[X]$, the inequality becomes $E[u(w+X-E[X])] \leq u(w)$, with equality if and only if X is degenerate. Thus an agent whose wealth is deterministic cannot improve its expected utility by purchasing at expected value. A non-degenerate stochastic cash flow must sell at less than its expected present value for the expected utility of the agent to increase. This is not always the case when wealth W is truly stochastic, since X might countervail W.

The q^* that solves the equation $E[u(W+X-q)] = u(W)$ satisfies the two principles of Appendix A. As per the first principle, it must lie within the minimum and maximum bounds of X. In particular, if X is degenerate, q^* must equal X with probability one. As per the second, for any real number k, $E[u(W+(X+k)-(q+k))] = u(W)$ if and only if $E[u(W+X-q)] = u(W)$. Therefore, a change in the level of X can be offset only by the same change in q.

The strengths of this version of utility theory notwithstanding, it has two defects that may explain why many sympathizers remain unconvinced. First, the comparison $E[u(W+X-q)] \sim u(W)$ allows one to calculate the unique q^* at which both sides are

equal, at which price the economic agent merely conserves its expected utility. But a ceiling or maximum price is not the same as a bidding price. Knowing not to pay more than 100 dollars for something is not quite the same as knowing what to pay for it. And second, these ceiling prices are not linear; for example, the ceiling price for twice X is not twice the ceiling price for X. The next section will remove these defects.

5. OPTIMAL UTILITY AND INDIVIDUAL EQUILIBRIUM

To determine whether $E[u(W+X-q)] > u(W)$ is to concentrate on the price q. Perhaps we should concentrate not on the price q at which to buy the whole X, but rather on the amount of X to buy at the price q. It may not be realistic to buy a fraction of a house, or a negative amount of land; but stochastic cash flows are ideally scaleable. The expected utility of an agent that purchases θ units of X at price q per unit is $f(\theta,q) = E[u(W+\theta X-\theta q)]$. The derivative with respect to q is $\frac{\partial f}{\partial q} = -\theta \cdot E[u'(W+\theta X-\theta q)]$. Since u' is positive, the expectation is positive, and the sign of the derivative is opposite to that of θ. If θ is positive, a decrease in price q increases expected utility f. And if θ is negative, an increase in price q increases f. Hence, a buyer (with positive θ) seeks as low a price as possible, whereas a seller (with negative θ) seeks as high a price as possible. Nothing more than this obvious truth is to be gained from concentrating on price.

A more subtle and fruitful perspective is to treat q as given and to determine how much of X to purchase. The derivative of the expected utility with respect to θ is

$\frac{\partial f}{\partial \theta} = E[u'(W + \theta X - \theta q)(X - q)]$. Appendix B proves that if $P(X > q)$ and $P(X < q)$ are both greater than zero (i.e., that $X - q$ has both upside and downside potential), then f as a function of θ looks like a concave downward parabola. In symbols, $\lim_{\theta \to \pm\infty} f(\theta, q) = -\infty$. And f strictly increases up to a maximum, after which it strictly decreases. Hence, an economic agent that has little influence on price turns its attention to buying the unique amount that optimizes its expected utility. What determines price is left to Section 6; but for now one might reasonably expect prices to be less erratic and more reflective of value when the concern of agents is how much to buy or to sell, rather than to negotiate the most advantageous deals.

Exhibit 5 graphs $f(\theta)$ for various values of q. In this example $W = 0$, X is a fifty-percent chance of receiving 100 dollars, and the agent's utility is exponential, viz., $u(w) = -e^{-0.1w}$. Not unexpectedly, all the curves intersect at $\theta = 0$, for the price of X is irrelevant if the agent has none; in symbols, $f(0, q) = E[u(W)] = E[-e^0] = -1$. As long as $0 < q < 100$ the curves rise from negative infinity, attain a maximum, and fall back to negative infinity. If $0 < q < E[X] = 50$, the risk-averse agent will optimize its expected utility with a long position ($\theta > 0$). If $E[X] < q < 100$, a short position ($\theta < 0$) is maximal; at $q = 50$ the agent neither buys nor sells. If q were 0 (a "can't lose" situation), or even negative (a "must win" situation), the agent would buy an infinite amount of X. Similarly, if q were 100 dollars (a "can't win" situation) or greater (a "must lose" situation) the agent would sell an infinite amount. In addition to four curves with maxima, the graph shows curves for the "must lose" situation of $q = 110$ and for the "can't lose" situation of $q = 0$. Due to

the fact that this agent cannot distinguish X from $100 - X$, the family of curves is symmetric about the vertical line $\theta = 0$, i.e., $f(-\theta, 100 - q) = f(\theta, q)$.

The total differential of f is:

$$\begin{aligned} df &= \frac{\partial f}{\partial \theta} d\theta + \frac{\partial f}{\partial q} dq \\ &= E[u'(W + \theta X - \theta q)(X - q)]d\theta - \theta E[u'(W + \theta X - \theta q)]dq \end{aligned}$$

Since u' is everywhere positive, $E[u'(W + \theta X - \theta q)] > 0$. Therefore, an increase in price (a positive dq) causes a change in f whose sign is opposite to that of θ. In other words, on the positive or long side of the θ-axis high-q curves lie below low-q curves. Correspondingly, on the negative or short side of the θ-axis high-q curves lie above low-q curves. As one can see in the exhibit, to increase price causes the expected utility of long positions to decrease and the expected utility of short positions to increase

From this point we restrict our attention to those values of q for which both Prob[$X < q$] and Prob[$X > q$] are greater than zero, for which values $f(\theta)$ has a maximum. Amount θ at price q is optimal if and only if:

$$\frac{\partial f}{\partial \theta} = E[u'(W + \theta X - \theta q)(X - q)] = 0$$

One would like to know how the optimal amount varies with price; in particular, one would expect it to vary inversely with price. The total differential of this partial derivative is:

$$\begin{aligned} d\left(\frac{\partial f}{\partial \theta}\right) &= \frac{\partial E[u'(W + \theta X - \theta q)(X - q)]}{\partial \theta} d\theta + \frac{\partial E[u'(W + \theta X - \theta q)(X - q)]}{\partial q} dq \\ &= E[u''(W + \theta X - \theta q)(X - q)^2]d\theta - E[u''(W + \theta X - \theta q)\theta(X - q) + u'(W + \theta X - \theta q)]dq \end{aligned}$$

Optimal ($d\theta$, dq) combinations maintain the total differential at zero:

$$0 = E\left[u''(W+\theta X-\theta q)(X-q)^2\right]d\theta - E\left[u''(W+\theta X-\theta q)\theta(X-q)+u'(W+\theta X-\theta q)\right]dq$$

Because u'' is everywhere negative the coefficient of $d\theta$ is negative. And because u' is everywhere positive, the second term within the coefficient of dq is positive. If the whole coefficient of dq is positive, $d\theta$ and dq will vary inversely. It may not be positive in all cases, even with the fact that $E[u'(W+\theta X-\theta q)(X-q)]=0$. But in the case of exponential utility, $u''(w)=-a^2e^{-aw}=-a\cdot ae^{-aw}=-a\cdot u'(w)$. Consequently:

$$\begin{aligned} E[u''(W+\theta X-\theta q)\theta(X-q)] &= E[-a\cdot u'(W+\theta X-\theta q)\theta(X-q)] \\ &= (-a\theta)E[u'(W+\theta X-\theta q)(X-q)] \\ &= (-a\theta)\cdot 0 \\ &= 0 \end{aligned}$$

Then the whole coefficient of dq is positive. In behavioral terms, with an increase of q the agent deems a decrease of θ to be optimal. But even without exponential utility, for optimal θ to vary inversely with q should be usual; real agents buy less at higher prices, and more at lower.

An economic agent can optimize its utility simultaneously for many stochastic cash flows. Let the present values of the cash flows be X_j for j = 1, 2, ... , and let them be offered at prices q_j. So the agent seeks to maximize $f(\Theta)=E\left[u\left(W+\sum_j\theta_jX_j-\sum_j\theta_jq_j\right)\right]$.

The partial derivatives $\frac{\partial f}{\partial\theta_k}=E\left[u'\left(W+\sum_j\theta_jX_j-\sum_j\theta_jq_j\right)(X_k-q_k)\right]$ must be zero at the optimal values θ^*. Letting W^* denote the optimal portfolio $W+\sum_j\theta_j^*X_j-\sum_j\theta_j^*q_j$, we must have for all k, $E[u'(W^*)(X_k-q_k)]=0$. This means that the prices of the stochastic

cash flows with respect to which an agent has optimized its expected utility, even though the agent passively accepted them, satisfy the equations:

$$q_j = \frac{E[u'(W^*)X_j]}{E[u'(W^*)]} = E[\Psi X_j]$$

A portfolio optimized with respect to cash flows X_j is also in equilibrium with respect to them; in other words, the agent wishes neither to buy nor to sell amounts of X_j. The agent can express the value of each cash flow as $q_j = E[\Psi X_j]$. This may not be true of every stochastic cash flow; but it is true of the flows with respect to which the portfolio is optimized.

The random variable Ψ is equal to $u'(W^*)/E[u'(W^*)]$, and $E[\Psi]=1$. Ψ is called variously a state-price random variable (Halliwell [2001] and Panjer [1998]), a deflator (Christofides and Smith [2001]), and a Radon-Nikodym derivative (Gerber and Pafumi [1998]). In terms of the probability system, Ψ maps from elementary outcome ξ_i to:

$$\frac{u'(W^*(\xi_i))}{\sum_j P(\xi_j)\cdot u'(W^*(\xi_j))}$$

Because of diminishing marginal utility u' will be less positive for the favorable outcomes of W^* than it will be for the unfavorable. Therefore $E[\Psi X]=\sum_i P(\xi_i)\Psi(\xi_i)X(\xi_i)$ reweights the present value of a stochastic cash flow, giving more weight to the outcomes at which the portfolio is less valuable.[11] Hence, if X tends to vary with W^*, $E[\Psi X]$ will tend to be

[11] So $P(\xi)\cdot\Psi(\xi)$ qualifies as a probability measure (Section 2). That it equals zero if and only if $P(\xi) = 0$ makes it an equivalent measure. Wang [2001] refers to $P(\xi)\cdot\Psi(\xi)$ as distorted probability.

less than $E[X]$. If it tends to vary against W^*, $E[\Psi X]$ will tend to be greater.

A perfectly optimized portfolio would be in equilibrium with respect to all stochastic cash flows. Equipoised and wishing to buy or to sell nothing, its economic agent would value every cash flow as $q_X = E[u'(W^*)X]/E[u'(W^*)] = E[\Psi X]$. This formula satisfies the principles of Appendix A. For first, it handles correctly the degenerate cash flow:

$$q_k = E[\Psi k] = E[\Psi] \cdot k = 1 \cdot k = k$$

And second, as a weighted average it stays within any bounds. Moreover, it is linear:

$$q_{\sum \alpha_i X_i} = E[\Psi(\sum \alpha_i X_i)] = \sum \alpha_i E[\Psi X_i] = \sum \alpha_i q_{X_i}$$

The linearity of valuation is paradoxical. Valuation should be linear, and indeed is linear, if an agent is in equilibrium. However, if the agent is in equilibrium, it neither buys nor sells. But an agent that buys or sells believes that the result will restore equilibrium, and buys or sells into the equilibrium price. In other words, the valuation of economic agents anticipates equilibrium.

6. PRICE DETERMINATION AND MARKET EQUILIBRIUM

Until now we have thought of economic agents as having no influence on price. But consider a market of $n > 1$ agents.[12] Since at any given moment stochastic cash flow C, with present value X, can have just one price, the i^{th} agent seeks to maximize its expected utility $f_i(\theta_i) = E[u_i(W_i + \theta_i X_{(i)} - \theta_i q)]$, the optimal amount being θ_i^*. In this

[12] There is no need to distinguish between large and small markets (as per Wang [2001, §2]), a distinction that continuity of size renders arbitrary.

formulation each agent has its own utility function, current wealth, and optimum. Moreover, each has its own estimate of the properties of X, which is indicated by a subscript in parentheses, $X_{(i)}$. But really this means that each agent has its own probability system. The only quantity common to all is the price q. Sellers belong as much as buyers to the market, selling being just the buying of a negative amount and vice versa.

Whether an agent regards itself as a price-maker, a price-taker, or something in between, for any given q, the agent should purchase the amount θ that maximizes its expected utility. But in a market they collaborate by setting q such that their optimal θ^* amounts clear. Normally to clear is for net buying to be zero, i.e., $\sum_i \theta_i^* = 0$. However, it could be otherwise, e.g., $\sum_i \theta_i^* = 1$, as in an auction. There is one and only one price q at which optimal amounts θ^* clear. For at higher prices, the expected utilities are maximized at lesser θ values, which means that sellers want to sell more than buyers want to buy. Conversely, at lower prices, the utilities are maximized at greater θ values, and buyers want to buy more than sellers want to sell.

In mathematical terms, a market of n agents is a mechanism for solving a system of n+1 equations in n+1 variables. The n+1 equations consist of the n equations that maximize the expected utilities (or that set their first derivatives to zero) and the clearing equation $\sum_{i=1}^{n} \theta_i^* = 0 \ (\text{or } 1)$. And the n+1 variables are the n values of θ and the one price q.

Essentially, price is a device whereby the agents of a market maximize their expected utilities. Most agents claim to have little or no influence on price, and great or total influence on quantity. But if the market solves a system of simultaneous equations, then in overall equilibrium the random wealth W^* of each agent is more than acceptable; as far as cash flow X is concerned it is optimal. The agents collaborate, mostly unawares, to arrive at an equilibrium in which each does as well as possible.[13]

The remaining exhibits develop this idea in a three-agent market. The three agents are A, B, and C, or Abel, Baker, and Charlie for realism. At first, let us assume that the wealth of each is deterministic at 100,000 dollars, and that their utility functions are the same at $u(x) \propto -e^{-0.000005x}$. An agent at this wealth and with this utility would regard a gain of 11,000 dollars and a loss of 10,000 dollars as offsetting.

Exhibit 6 describes how these agents would price stochastic cash flow X, which they all believe to be a fifty-percent chance of gaining 100,000 dollars. Four parameters determine the exhibit, one price q and three amounts θ, one for each agent. These parameters are the shaded cells of the topmost table of the exhibit. However, the θ amount for Charlie is constrained so that the three amounts total to one. In other words, stochastic cash flow X must be completely auctioned off to Abel, Baker, and Charlie. So really, the free parameters are the unit price q and the θ amounts of Abel and Baker.

[13] Such an overall equilibrium is known as a Pareto optimum (Gerber, 1979, Chapter 7), technically defined as a state in which one cannot do better unless another does worse.

The next table simply lists the marginal utilities of the agents, which here are the same. The next three tables take q and the θ amounts and calculate each agent's expected marginal utility, i.e., $E[u'(W+\theta X-\theta q)(X-q)]$. This appears in the last cell of the "Mean" row of each agent's table, which is the sumproduct of the columns labeled "Probability" and "$u'(\ )(X\text{-}q)$". The expected marginal utilities are fed into the last column of the topmost table, and their summary statistic in the "Total" row is their root mean square, i.e., the square root of the average of their squares.

The Excel Solver add-in seeks q, Abel's θ, and Baker's θ so as to minimize the root mean square. According to theory, there is one, and only one, equilibrium at which all three expected marginal utilities are zero, which happens if and only if their root mean square is zero. The solver deemed 5.959E-03 as the minimum, with individual marginal utilities 1.931E-03, 9.495E-03, 3.555E-03. Being six or seven orders of magnitude less than the state marginal utilities, these amounts are effectively zero. So Abel, Baker, and Charlie buy equal shares (33.33%) of X, which makes sense since they have the same utility and the same outlook as to Probability, W, and X. And together they buy all of X at a price of 45,843 dollars. At a higher price they would not buy it all, and at a lower price they would want to buy too much. It doesn't matter here from where X enters the financial universe of Abel, Baker, and Charlie. There does not have to be a fourth agent who owns all of X. One may posit an infinitely risk-averse quasi-agent, such as "Luck" or "Possibility," which will unload X at any price with upside potential.

In Exhibit 7 Abel is less risk-averse than Baker and Charlie; his exponential parameter $a = 2.500\text{E}-06$ is half theirs. They all have the same outlook; but Abel has twice the

appetite for risk. This implies that the price must rise from 45,843 to discourage overbuying. The solver arrived at the price of 46,879 dollars, at which price Abel takes half and Baker and Charlie take quarters. Abel, in effect, functions like two agents: If a fourth agent were added to Exhibit 6 having the same parameters and outlook as Abel, Baker, and Charlie, the solver would arrive at the price of 46,879 and each would have a quarter share.

Exhibit 8 takes a step into the reality of differing opinions. Charlie continues to believe that X is a fifty-percent chance of gaining 100,000 dollars. But Abel assesses it as a fifty-percent chance of gaining 80,000, and Baker as a forty-percent chance of gaining 100,000. All have the same utility; but to Abel and Baker stochastic cash flow X is less attractive than it was in Exhibit 6. Therefore, it is no surprise that the equilibrium comes at the lower price of 39,353 dollars, at which Abel's and Baker's shares are only 8.09% and 5.41%. Charlie rates X highly, and buys 86.50% of it. This example shows that the market model developed in this paper does not require the economic agents to have the same assessment of risk. Each agent must be free to create its own probability system.

Finally, Exhibit 9 treats the spreading of risk within a closed system. Charlie owns X (in addition to his deterministic wealth of 100,000 dollars). And Charlie's θ amount is constrained so that the three amounts total to zero. Otherwise, the exhibit is the same as Exhibit 6. Not unexpectedly, the equilibrium is the same; the price is 45,843 dollars, and each agent ends up with a third of X.

This market-tempered utility theory generalizes to any number of agents and any risk-

averse utility functions. Moreover, the probability system of each agent, in which its wealth W and the cash flow X are related, is arbitrary. Each agent is free to act on its own beliefs, whether accurate or not. Even as to the issue of present value, mentioned in Section 2, each agent is free to make its own judgments. But this model countenances no difference between large and small markets; it holds as much for a market of two agents as for a market of millions. A market would seem to be no more and no less than the sum of its parts, i.e., a number of agents each seeking to maximize its own expected utility. It is agents that move their markets, not markets their agents.

7. REINSURANCE PRICING

It is instructive to apply this valuation theory to reinsurance. Suppose that an insurance company asks reinsurers to assume a catastrophic risk, e.g., its exposure to hurricane losses in excess of a certain threshold. We will assume that the insurer wants, or is constrained, to reinsure this risk at any reasonable cost. Exhibit 10 shows how a market of two reinsurers might determine the price. The exhibit is set up for insurance losses L and premiums p, rather than for payoffs X and prices q. But Section 4 explained how losses are equivalent to negative payoffs, and premiums to negative prices.

Reinsurer A is freshly capitalized at one billion dollars, which it has placed in cash and government securities. After examining the underwriting information, it estimates the probability of losing fifty million dollars to be four percent, and that of losing one hundred million to be two percent. So the expected loss is four million dollars. Reinsurer B already has an insurance and investment portfolio; in fact, it estimates the distribution of

the present value of its wealth as twenty-five percent at 1.1 billion dollars, fifty percent at 1.0 billion, and twenty-five percent at 0.9 billion. So its size is comparable to that of Reinsurer A, and we've assumed both their utilities to be exponential with *a*-parameter 5.000E-09.

Though Reinsurer B agrees with Reinsurer A that the expected loss is four million dollars, it believes the probability of losing twenty-five million dollars to be one percent, that of losing fifty million dollars to be two and a half percent, and that of losing one hundred million to be two and a half percent. The "Probability" column of Reinsurer B in the exhibit is the outer product of (25%, 50%, and 25%) and (94%, 4%, 2%). If loss column "*L*" consisted of three blocks of (0, 50, 100) million dollars, the hurricane loss would be independent of the stochastic wealth of Reinsurer B. If this were the case, exponential utility would allow Reinsurer B to value this risk on its own (Appendix C). Both reinsurers would have the same assessment of the risk, and each would take a fifty percent share at the total premium of 4,781,718 dollars.[14]

However, Reinsurer B does not deem the risk as independent; the tendency is for the loss to be greater when the current stochastic wealth is less. In fact, the correlation coefficient between *W* and *L* is –8.2 percent. Perhaps Reinsurer B has already assumed some of the same hurricane risk from other insurers. In any case, it is not as eager to assume the new risk as is Reinsurer A. If fact, a total premium of 5,501,545 dollars is needed for a fifty percent share of the risk to maximize the expected utility of

[14] Each reinsurer receives its share of the premium. The θ factors in the fifth and sixth columns of each reinsurer's calculation take this into account.

Reinsurer B, versus the 4,781,718 for Reinsurer A. The undesirable covariance increases the quote of Reinsurer B by more than 700,000 dollars. But the reinsurers together will assume the whole risk for a total premium of 5,106,791 dollars, Reinsurer A signing the slip for 68.29 percent and Reinsurer B for 31.71 percent.

One often hears about erratic swings in the reinsurance market. Exhibit 11 suggests how this valuation theory might help to moderate such swings. Suppose that there is a third reinsurer, Reinsurer C, freshly capitalized just like Reinsurer A. But this reinsurer is dour about the hurricane risk, estimating a ten percent chance of losing fifty million dollars, and a three percent chance of losing one hundred million. So it expects eight million dollars of loss, twice as much as what the others expect. Therefore, it sits on the sidelines, lamenting the overcapitalized and soft reinsurance market while Reinsurers A and B offer 5.1 million dollars. Reinsurer C would be content with an equal share of the risk, but a total premium of 8.9 million dollars would be needed to maximize its expected utility at a one-third share (and 10.8 million at a full share).

Why not let Reinsurer C sign the placement slip for a negative share? This is the equivalent to short selling in other markets; so why not here? Implicitly the insurance company is on the slip at a –100 percent share. The short positions on the slip can lose at most the premiums that they pay; it is the long ones that are "naked" to the losses. In fact, since losses and premiums are inverses of payoffs and prices, it is hard to define here what is long and what is short. According to the exhibit, Reinsurer C expresses its belief in cheap reinsurance by taking a –90.91 percent share. The effect of its contrariness is to raise the price from 5,106,791 to 6,077,896 dollars. Instantly the

market hardens by nineteen percent, and everyone (other than, perhaps, the insurance company) is happy. One might have qualms about a reinsurer's having more than a 100 percent share; but this is less likely to happen in a market of more than three.

Though it is not uncommon for direct-writing reinsurers to cooperate and to take shares of risks, reinsurance brokers would have the advantage to implement this particular strategy of negative shares, even as they have advantages to foster cooperation among reinsurers. However, at present reinsurers are not prepared for market pricing, for few understand that the premium for a share of a deal should not equal the share of the premium for the whole deal.[15]

8. IMPLICATIONS

Before concluding, we will here draw out implications for two important and relevant topics, solvency and investment. As to solvency, risk-adjusted discounting has misled many to elevate it from the status of a constraint to that of a valuation method. Solvency is important, especially for insurance and reinsurance companies; so important, in fact, that they ought to decline deals that jeopardize it no matter how attractively priced they are. But the insurance industry, along with its regulators and analysts, has adopted the method of the banking industry, as expressed in the so-called "Basle Accord" (Basle Committee, 1988). This method is concerned with how a company's net worth might deteriorate from one (annual) balance sheet to the next. Capital is charged against, or

[15] Again, risk-adjusted discounting is the culprit. Rates of return are scaleable; the ROE of half the deal at half the price is the same as the ROE of the whole deal at the whole price.

allocated to, elements of wealth that are subject to various risks. A comparison of net worth with this overall "risk-based capital" determines whether the company is financially strong or weak.

This has led many, both in banking and insurance, to cost-of-capital pricing. One calculates how much capital to allocate to a deal and subtracts the cost of this capital from the deal's expected value. Now, to be fair, the solvency regulations refrain from saying how to price banking and insurance products. But lack of vice is not virtue. For regulators do believe cost-of-capital pricing to be a proper application of their solvency regulations. In fact, they can't imagine how pricing could be done otherwise, and they are gratified to see their industries following in their trains. This busies many technical minds with constructing ideal risk measures, according to which one may allocate enough, but not too much, capital to risky deals. As Glenn Meyers [2002] expresses it, "We have to balance the cost of an insolvency with the cost of holding capital."

Both the regulation and the application are flawed, the flaw in both concerning time. First, the regulation countenances only accounting items and how they might change from statement to statement. But cash, not capital, is the *prima materia* of the financial universe. To stress the current balance sheet with an array of one-year assaults tells something about financial strength; but it does not plumb to the depth of cash. Better is to estimate the probability for a company's cash account to remain positive throughout its runoff. A company highly certain to have enough cash on hand to quit its obligations should be deemed stronger than a company highly certain to have a positive net worth one year hence. This superior criterion of strength requires peering into the cash level.

Second, nearly everyone blithely assumes that money works, in particular, that it works at a certain rate per time period. So one hundred dollars working at ten percent per year should be paid ten dollars at the end of the year.[16] But if a risk were resolved in a very short time, days or even hours, capital allocated to it should be paid pro-rata as to time. Unlike repairmen, who charge by the hour or fraction thereof, capital does not bill by the whole year.[17] The whole picture of insurance needs revamping. Fussing over regulation and its effect upon accounting statements evades economics.[18] Insurance is not the exposing of capital to loss and the pooling of risks to optimize such exposure. Rather, it is a transaction in which a stochastic cash flow is exchanged to the benefit of the expected utilities of both parties (Bowers, 1986, Chapter 1, aptly titled "The *Economics* of Insurance").

As to the second topic, investment is a species of valuing stochastic cash flows. An insurance or reinsurance company ought always to revise its estimate of the present value of the sum of its stochastic cash flows. To keep negligible the probability that this present-value random variable is negative (perhaps more accurately, the probability that its run-off cash account would ever be negative) the company must limit both its insurance and investment activities. Investment neither piggybacks upon nor

[16] But who gets paid, the money or its owner? If the money works for its owner, is it a slave? Appendix A argues that for money to work is nonsensical and misleading language. Obviously it is a figure of speech. But if none bothers (or is able) to translate it into literal language, how can the figure be harmless?

[17] "The is no natural unit of time." (Halliwell, 1999, Appendix E) In the same section he teases out the ramifications to capital allocation of doing business on other planets. Venus might become the Bermuda of the next century, for its shorter year may afford reinsurers accounting advantages!

[18] One can improve wealth on paper without improving real wealth. The fault is not just with the present accounting rules; it will always exist, even after the eagerly awaited convergence of IAS and GAAP in 2005. One should make sound economic decisions and let the accounting chips fall where they may.

supercharges underwriting. Since each crowds out the other, the business of insurance should be to underwrite well, not to underwrite to generate funds to invest well.

Insurance deals are perfect specimens of stochastic cash flows; insurance is the ideal setting in which to apply this theory.[19] Investing in equities is far more complicated. For the stock investor puts money not into a project, but into a corporation whose employees will put money into projects indefinitely. Since one bets not on a deal, but on persons who will bet on deals, the equity investor is at one remove from the stochastic cash flows. It is not just a matter of buying in, possibly receiving a few dividend checks, and selling out; it is not even as simple as projecting a dividend stream, as per the Gordon dividend-growth model. To value a stock is to value the human management of a stochastic perpetuity. (No wonder that investment theory eschews stochastic cash flows!) In companies that understand this theory and the near idealness of its application to insurance chief actuaries will be kings; chief financial officers will be charged with corporate reporting and cash management. More than forty years ago Karl Borch predicted the ascendance of actuaries: "The traditional approach implies that the actuary should play a rather modest part in the management of his company. ... In the light of these theories [for decision making under uncertainty] it appears that the actuary should take a broader view of his duties." [1961, pp. 245f.]

---

[19] More accurately, it will be the ideal setting, if insurers and reinsurers should get back to basics, one of which is to underwrite exposures free of moral and morale hazards. According to Rob Jones of Standard & Poor's, "Reinsurers realize that they must focus on underwriting performance rather than investment portfolios." ("Reinsurers Must Get Back To Basics," *Reactions Rendez Vous Reporter*, September 11, 2002, p. 12, www.rvs-monte-carlo.com/docs/reaction/Reaction 11-09-2002.pdf.

9. CONCLUSION

From taking seriously the meaning of 'present' in 'present value' our theory of valuation derives. Present value really is value now, not value on the horizon of time, not value as of the next accounting statement. The present is right this moment, not after the next news flash, not even after a double take. Anything else, however cleverly concealed, is value in the future, to which the present is linked by capital working at percent per year. How strange that this understanding of 'present' is as countercultural to 1990s finance as such slogans as "Live for the moment!" were to 1960s society! But perhaps it will catch on in this decade.

The belief that value now must impound what it might later become confuses subject and object. "The price of an asset or a deal is different from the cash flow – as different as subject is from object." (Halliwell, 2001, §6) Hence, many cannot tell the difference between keen appraisers and savvy traders. Theorists construct models in which present price depends on future price, the more knowable on the less knowable. Predicting the future replaces valuing the present. This misconception is twin to that of subsuming the value of risk into the value of time. Recognizing this may well cause a financial revolution, as Oakley Van Slyke (1995, p. 587) writes:

> The theory of finance, both as it is and as it will be after the coming revolution, is by its nature prescriptive. The theory of finance suggests how a decision-maker should make decisions among alternative courses of action. The theory of finance is not descriptive. It does not show why people or institutions behave as they actually do.[20]

[20] Similarly Borch [1961, §5.2]: "Shackle does not consider his theory as normative in the sense that it states how rational businessmen should make decisions. All he claims is that his theory describes, or explains how businessmen actually reach their decisions." Which is more important, a description of how something is done or a description of how it ought to be done? Rationality has to do with oughtness.

One can enumerate risks of all sorts, e.g., that payoffs will be too small or too large, or too soon or too late, that interest rates will go up or down, that courts will be too lax or too stern, that the company comptroller will embezzle millions of dollars, that a virus will infect the company database, that an asteroid will strike the earth. These may or may not be legitimate concerns for the management of an insurance or reinsurance company. But valuation impounds into the price of a stochastic cash flow the distribution of the probability of its present value. Hence, valuation countenances only the risks that affect that distribution.

Present value is a random variable. An economic agent must know the present value W of its current stochastic cash flow. Then it must estimate the present value X of the proposed deal (e.g., as in Exhibit 2), as well as its joint distribution with W. Utility theory is the bridge from random variables to values. An agent should know how much θ of the deal to purchase at price q, i.e., the amount that maximizes its expected utility $f(\theta) = E[u(W + \theta X - \theta q)]$. Interaction of agents with one another will produce a clearing price, at which each agent purchases an optimal amount. The analytics of all this may seem too complex and difficult a task. However, insurance cash flows are perfect specimens of stochastic cash flows, and actuaries are fortunate to work in such an ideal laboratory. Furthermore, capital-allocation programs have grown at least as complex and difficult. Perhaps actuaries disenchanted with them will examine present value and utility, and give the finance world something to emulate.

References

Basel Committee on Banking Supervision, *International Convergence of Capital Measurement and Capital Standards*, Basle, 1988, www.bis.org/pub/bcbs04a.pdf.

Borch, Karl H., "The Utility Concept Applied to the Theory of Insurance," *ASTIN Bulletin*, 1(5), 1961, 245-255.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., and Nesbitt, C. J., *Actuarial Mathematics*, Society of Actuaries, 1986.

Brealey, Richard A., and Myers, Stewart, *Principles of Corporate Finance*, Seventh Edition, New York, McGraw-Hill, 2002.

Bühlmann, Hans, "An Economic Premium Principle," *ASTIN Bulletin*, 11(1), 1980, 52-60.

Butsic, Robert, "Determining the Proper Interest Rate for Loss Reserve Discounting: An Economic Approach," *Evaluating Insurance Company Liabilities*, 1988 CAS *Discussion Paper Program*, 147-188.

Christofides, Stauros, and Smith, Andrew D., "DFA – The Value of Risk," *CAS Forum Spring 2001: Including the Dynamic Financial Analysis Call Papers*, 153-194.

D'Arcy, Stephen P., "Introduction to the Discounted Cash Flow Approach," *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 19-25.

Daykin, C. D., Pentikäinen, T., and Pesonen, M., *Practical Risk Theory for Actuaries*, London, Chapman & Hall, 1994.

Debreu, Gerard, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New Haven, Yale, 1987.

Dixit, Avinash K., *Optimization in Economic Theory*, Second Edition, Oxford, 1990.

Duffie, Darrell, *Dynamic Asset Pricing Theory*, Second Edition, Princeton, 1996.

Gerber, Hans U., *An Introduction to Mathematical Risk Theory*, Philadelphia, S. S. Huebner Foundation, 1979.

Gerber, Hans U., and Pafumi, Gérard, "Utility Functions: From Risk Theory To Finance," *North American Actuarial Journal*, Volume 2 (July 1998), 74-100.

Halliwell, Leigh J., "ROE, Utility, and the Pricing of Risk," *1999 CAS Spring Reinsurance Call Papers*, 71-136.

Halliwell, Leigh J., "A Critique of Risk-Adjusted Discounting," *2001 ASTIN Colloquium*, www.casact.org/coneduc/reinsure/astin/2000/halliwell1.doc.

Kozik, Thomas J., and Larson, Aaron M., "The *n*-Moment Insurance CAPM," *PCAS* LXXXVIII (2001), 39-63.

Kreps, Rodney E., "Reinsurer Risk Loads from Marginal Surplus Requirements," *PCAS* LXXVII (1990), 196-203.

Lee, Yoong-Sin, "The Mathematics of Excess of Loss Coverages and Retrospective Rating – A Graphical Approach," *PCAS* LXXV (1988), 49-64.

Longley-Cook, Alastair G., "Risk-Adjusted Economic Value Analysis," *North American Actuarial Journal*, Volume 2 (Jan 1998), 87-100.

Mango, Donald, "Capital Consumption: An Alternative Methodology for Pricing Reinsurance," *CAS Forum Winter 2003*, 351-378.

Meyers, Glenn G., "Setting Capital Requirements with Coherent Measures of Risk – Part 1," *The Actuarial Review*, Volume 29 (August 2002), 13.

Panjer, Harry H., editor, *Financial Economics: with Applications to Investments, Insurance and Pensions*, Schaumburg, IL, Actuarial Foundation, 1998.

Preiffer, Paul E., *Concepts of Probability Theory*, New York, Dover, 1978.

Schnapp, Frank, "The Diversification Property," *Presented to the 11th AFIR Colloquium*, September 2001, www.actuaries.ca/meetings/afir/handouts/.

Sundt, Bjørn, *An Introduction to Non-Life Insurance Mathematics*, 2. Auflage, Karlsruhe, Verlag Versicherungswirtschaft, 1991.

Van Slyke, Oakley E., "A General Theory of Finance," *Proceedings of the 5th AFIR Colloquium*, September 1995, vol. II, 585-635.

Van Slyke, Oakley E., "The Cost of Capital: An Axiomatic Approach," *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 135-164.

Von Neumann, John, and Morgenstern, Oskar, *The Theory of Games and Economic Behavior*, Princeton, 1972.

Wang, Shaun S., "Equilibrium Pricing Transforms: New Results of Buhlmann's 1980 Economic Model," 2001, www.stats.uwaterloo.ca/Stats_Dept/IIPR/2001-reports/IIPR-01-19.pdf.

Exhibit 1

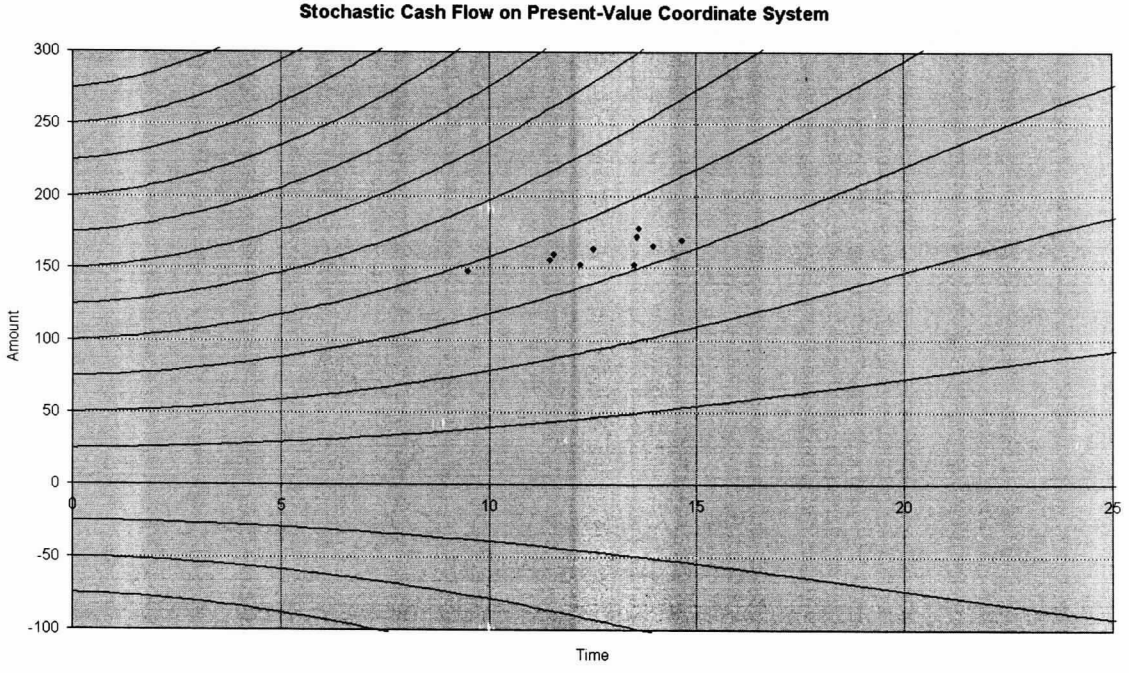

Stochastic Cash Flow on Present-Value Coordinate System
300
250
200
150
100
50
0
-50
-100
Amount
0
5
10
15
20
25
Time

Exhibit 2

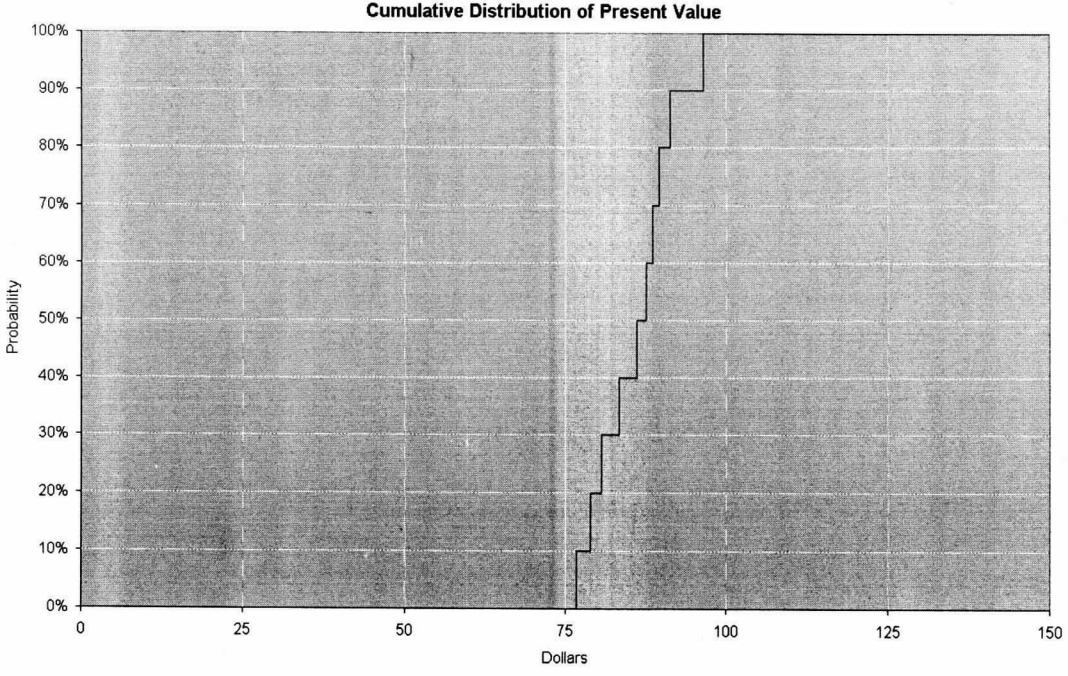

Cumulative Distribution of Present Value
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Probability
0
25
50
75
100
125
150
Dollars

Exhibit 3

Issuer	TYSON FOODS INC
Issue Date	01 Oct 2001
Maturity	01 Oct 2006
Price	747,187,500
Face Amount	750,000,000
Coupon	7.25%
Frequency	semi-annual
Moody's Rating	Baa3
Structure	senior unsecured

Recovery Parameters	
μ	44.62%
σ	26.32%
α	1.145
β	1.422
B(α, β)	40.00%
Recovery	300,000,000

		Without Default				Simulated Default				@ 01 October 2001	
t	Date	Default Probability	Actual	Principal	Interest	Default Probability	Actual	Principal	Interest	STRIP Prices	$v(t)$
0.0	01 Oct 2001	0.00%	1		0	0.00%	1		0	100 00/32	1.000
0.5	01 Apr 2002	0.00%	1		27,187,500	0.20%	1		27,187,500	98 27/32	0.988
1.0	01 Oct 2002	0.00%	1		27,187,500	0.31%	1		27,187,500	97 22/32	0.977
1.5	01 Apr 2003	0.00%	1		27,187,500	0.48%	1		27,187,500	96 02/32	0.960
2.0	01 Oct 2003	0.00%	1		27,187,500	0.75%	1		27,187,500	94 14/32	0.944
2.5	01 Apr 2004	0.00%	1		27,187,500	0.98%	1		27,187,500	92 18/32	0.925
3.0	01 Oct 2004	0.00%	1		27,187,500	1.28%	0		0	90 12/32	0.904
3.5	01 Apr 2005	0.00%	1		27,187,500	1.68%	0		0	88 06/32	0.882
4.0	01 Oct 2005	0.00%	1		27,187,500	2.21%	0		0	86 02/32	0.860
4.5	01 Apr 2006	0.00%	1		27,187,500	2.48%	0		0	84 07/32	0.842
5.0	01 Oct 2006	0.00%	1	750,000,000	27,187,500	2.79%	0	300,000,000	0	82 13/32	0.824
Present Value				618,046,875	247,614,404	1.00%		247,218,750	130,372,559		

Exhibit 4

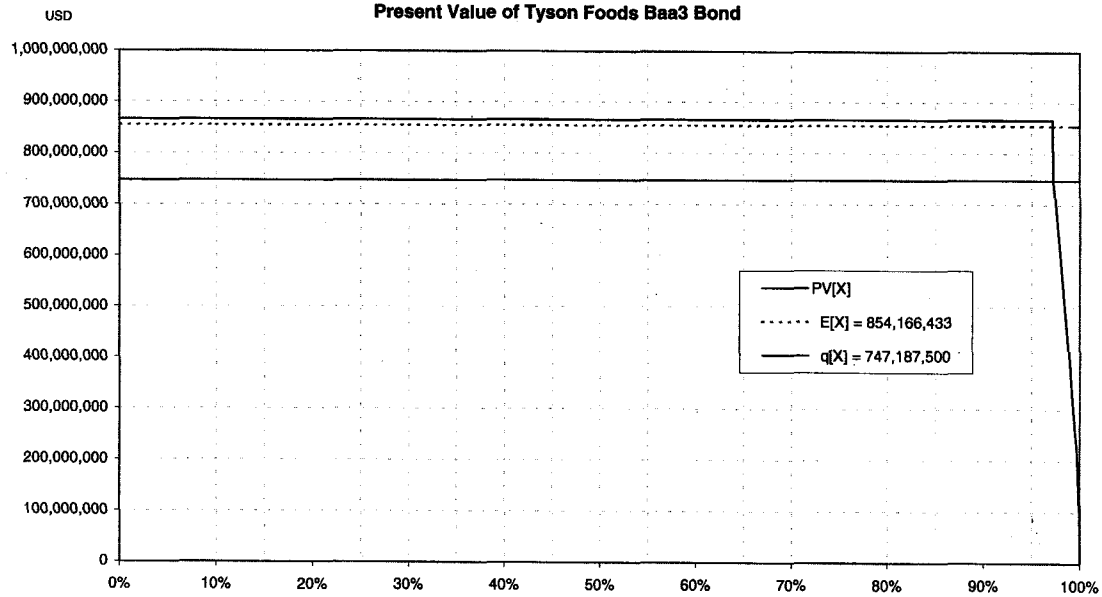

Present Value of Tyson Foods Baa3 Bond
USD
1,000,000,000
900,000,000
800,000,000
700,000,000
600,000,000
500,000,000
400,000,000
300,000,000
200,000,000
100,000,000
0
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
PV[X]
E[X] = 854,166,433
q[X] = 747,187,500

Exhibit 5

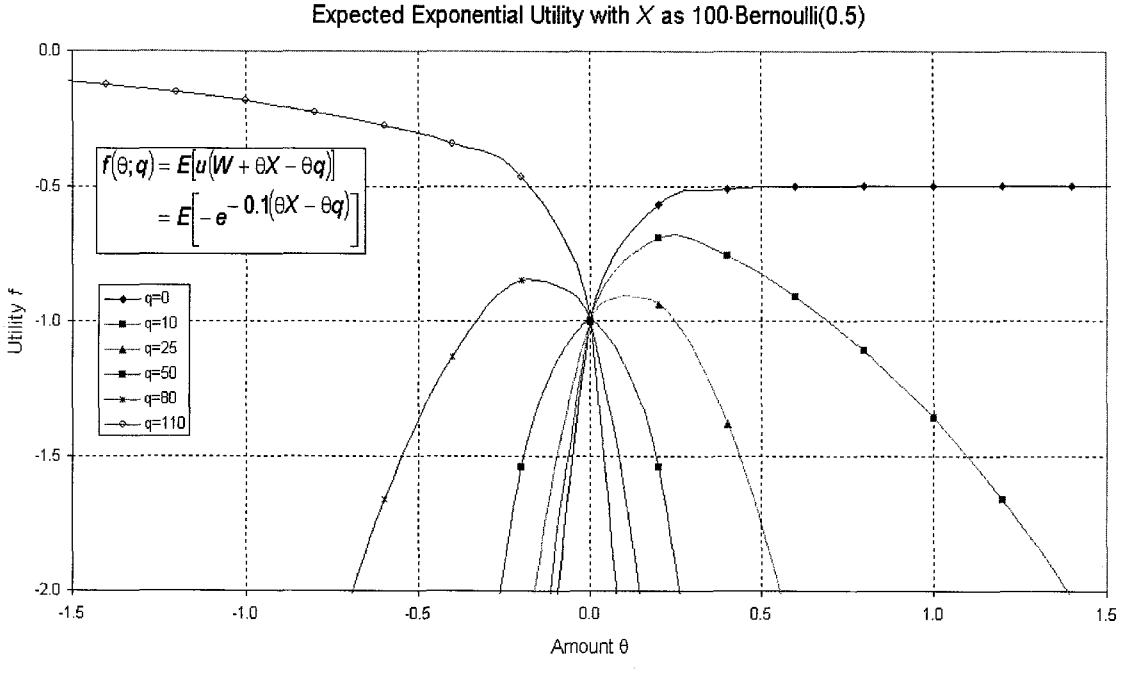

Expected Exponential Utility with X as 100·Bernoulli(0.5)
$f(\theta;q) = E[u(W + \theta X - \theta q)]$
$= E\left[-e^{-0.1(\theta X - \theta q)}\right]$
q=0
q=10
q=25
q=50
q=80
q=110
Utility f
Amount θ
0.0
-0.5
-1.0
-1.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

Exhibit 6

Three Identical Agents

Agent	Price q	θ	$E[u'(\,)(X-q)]$
Abel	45,843	33.33%	1.931E-03
Baker	45,843	33.33%	9.495E-03
Charlie	45,843	33.33%	3.555E-03
Total	45,843	100.00%	5.959E-03

Agent	a	$u'(x)$
Abel	5.000E-06	exp(-5.000E-06 * x)
Baker	5.000E-06	exp(-5.000E-06 * x)
Charlie	5.000E-06	exp(-5.000E-06 * x)

Abel	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-45,843	84,719	6.547E-01	-3.001E+04
State 2	50%	100,000	100,000	54,157	118,052	5.542E-01	3.001E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	4,157	101,386		1.931E-03

Baker	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-45,843	84,719	6.547E-01	-3.001E+04
State 2	50%	100,000	100,000	54,157	118,052	5.542E-01	3.001E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	4,157	101,386		9.495E-03

Charlie	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-45,843	84,719	6.547E-01	-3.001E+04
State 2	50%	100,000	100,000	54,157	118,052	5.542E-01	3.001E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	4,157	101,386		3.555E-03

Exhibit 7

One Agent Less Risk-Averse

Agent	Price q	θ	$E[u'(\,)(X-q)]$
Abel	46,879	50.00%	2.005E-03
Baker	46,879	25.00%	1.425E-02
Charlie	46,879	25.00%	1.500E-02
Total	46,879	100.00%	1.200E-02

Agent	a	$u'(x)$
Abel	2.500E-06	exp(-2.500E-06 * x)
Baker	5.000E-06	exp(-5.000E-06 * x)
Charlie	5.000E-06	exp(-5.000E-06 * x)

Abel	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-46,879	76,560	8.258E-01	-3.871E+04
State 2	50%	100,000	100,000	53,121	126,561	7.288E-01	3.871E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	3,121	101,560		2.005E-03

Baker	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-46,879	88,280	6.431E-01	-3.015E+04
State 2	50%	100,000	100,000	53,121	113,280	5.676E-01	3.015E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	3,121	100,780		1.425E-02

Charlie	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-46,879	88,280	6.431E-01	-3.015E+04
State 2	50%	100,000	100,000	53,121	113,280	5.676E-01	3.015E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	3,121	100,780		1.500E-02

Exhibit 8

Agents of Differing Opinions

Agent	Price q	θ	$E[u'(\,)(X-q)]$
Abel	39,353	8.09%	1.431E-02
Baker	39,353	5.41%	3.093E-03
Charlie	39,353	86.50%	6.819E-03
Total	39,353	100.00%	9.327E-03

Agent	a	$u'(x)$
Abel	5.000E-06	exp(-5.000E-06 * x)
Baker	5.000E-06	exp(-5.000E-06 * x)
Charlie	5.000E-06	exp(-5.000E-06 * x)

Abel	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-39,353	96,816	6.163E-01	-2.425E+04
State 2	50%	100,000	80,000	40,647	103,289	5.966E-01	2.425E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	40,000	647	100,052		1.431E-02

Baker	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	60%	100,000	0	-39,353	97,872	6.130E-01	-2.412E+04
State 2	40%	100,000	100,000	60,647	103,280	5.967E-01	3.619E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	40,000	647	100,035		3.093E-03

Charlie	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-39,353	65,959	7.191E-01	-2.830E+04
State 2	50%	100,000	100,000	60,647	152,461	4.666E-01	2.830E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	10,647	109,210		6.819E-03

Exhibit 9

Agents as a Secondary Market

Agent	Price q	θ	$E[u'(\,)(X-q)]$	Comment
Abel	45,843	33.33%	1.436E-02	Charlie owns X
Baker	45,843	33.33%	8.972E-03	
Charlie	45,843	33.67%	5.156E-03	
Total	45,843	0.00%	1.022E-02	

Agent	a	$u'(x)$
Abel	5.000E-06	exp(-5.000E-06 * x)
Baker	5.000E-06	exp(-5.000E-06 * x)
Charlie	5.000E-06	exp(-5.000E-06 * x)

Abel	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-45,843	84,719	6.547E-01	-3.001E+04
State 2	50%	100,000	100,000	54,157	118,052	5.542E-01	3.001E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	4,157	101,386		1.436E-02

Baker	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-45,843	84,719	6.547E-01	-3.001E+04
State 2	50%	100,000	100,000	54,157	118,052	5.542E-01	3.001E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	100,000	50,000	4,157	101,386		8.972E-03

Charlie	Probability	W	X	$X-q$	$W+\theta X-\theta q$	$u'(W+\theta X-\theta q)$	$u'(\,)(X-q)$
State 1	50%	100,000	0	-45,843	130,562	5.206E-01	-2.386E+04
State 2	50%	200,000	100,000	54,157	163,895	4.407E-01	2.386E+04
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	150,000	50,000	4,157	147,229		5.156E-03

Exhibit 10

Reinsurance Pricing

Agent	Premium p	θ	$E[u'(\,)(-L+p)]$
Reinsurer A	5,106,791	68.29%	4.276E-03
Reinsurer B	5,106,791	31.71%	-5.540E-03
Total	5,106,791	100.00%	4.948E-03

Agent	a	$u'(x)$
Reinsurer A	5.000E-09	exp(-5.000E-09 * x)
Reinsurer B	5.000E-09	exp(-5.000E-09 * x)

Reinsurer A	Probability	W	L	$-L+p$	$W-\theta L+\theta p$	$u'(W-\theta L+\theta p)$	$u'(\,)(-L+p)$
State 1	94.0%	1,000,000,000	0	5,106,791	1,003,487,660	6.621E-03	3.381E+04
State 2	4.0%	1,000,000,000	50,000,000	-44,893,209	969,340,385	7.854E-03	-3.526E+05
State 3	2.0%	1,000,000,000	100,000,000	-94,893,209	935,193,109	9.317E-03	-8.841E+05
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	1,000,000,000	4,000,000	1,106,791	1,000,755,878		4.276E-03

Reinsurer B	Probability	W	L	$-L+p$	$W-\theta L+\theta p$	$u'(W-\theta L+\theta p)$	$u'(\,)(-L+p)$
State 1	23.5%	1,100,000,000	0	5,106,791	1,101,619,131	4.054E-03	2.070E+04
State 2	1.0%	1,100,000,000	25,000,000	-19,893,209	1,093,692,769	4.218E-03	-8.390E+04
State 3	0.5%	1,100,000,000	50,000,000	-44,893,209	1,085,766,407	4.388E-03	-1.970E+05
State 4	47.0%	1,000,000,000	0	5,106,791	1,001,619,131	6.684E-03	3.413E+04
State 5	2.0%	1,000,000,000	50,000,000	-44,893,209	985,766,407	7.235E-03	-3.248E+05
State 6	1.0%	1,000,000,000	100,000,000	-94,893,209	969,913,682	7.832E-03	-7.432E+05
State 7	23.5%	900,000,000	0	5,106,791	901,619,131	1.102E-02	5.627E+04
State 8	1.0%	900,000,000	100,000,000	-94,893,209	869,913,682	1.291E-02	-1.225E+06
State 9	0.5%	900,000,000	100,000,000	-94,893,209	869,913,682	1.291E-02	-1.225E+06
State 10							
Mean	100%	1,000,000,000	4,000,000	1,106,791	1,000,350,913		-5.540E-03

	Probability	W	L	$-L+p$	$W-\theta L+\theta p$	$u'(W-\theta L+\theta p)$	$u'(\,)(-L+p)$
State 1							
State 2							
State 3							
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	0%	0	0	0	0		0.000E+00

Exhibit 11

Reinsurance Pricing with a Short Position

Agent	Premium p	θ	$E[u'(\,)(-L+p)]$
Reinsurer A	6,077,896	116.45%	-1.891E-02
Reinsurer B	6,077,896	74.45%	6.643E-04
Reinsurer C	6,077,896	-90.91%	-9.633E-03
Total	6,077,896	100.00%	1.226E-02

Agent	a	$u'(x)$
Reinsurer A	5.000E-09	exp(-5.000E-09 * x)
Reinsurer B	5.000E-09	exp(-5.000E-09 * x)
Reinsurer C	5.000E-09	exp(-5.000E-09 * x)

Reinsurer A	Probability	W	L	$-L+p$	$W-\theta L+\theta p$	$u'(W-\theta L+\theta p)$	$u'(\,)(-L+p)$
State 1	94.0%	1,000,000,000	0	6,077,896	1,007,077,996	6.504E-03	3.953E+04
State 2	4.0%	1,000,000,000	50,000,000	-43,922,104	948,850,639	8.702E-03	-3.822E+05
State 3	2.0%	1,000,000,000	100,000,000	-93,922,104	890,623,282	1.164E-02	-1.093E+06
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	1,000,000,000	4,000,000	2,077,896	1,002,419,808		-1.891E-02

Reinsurer B	Probability	W	L	$-L+p$	$W-\theta L+\theta p$	$u'(W-\theta L+\theta p)$	$u'(\,)(-L+p)$
State 1	23.5%	1,100,000,000	0	6,077,896	1,104,525,231	3.995E-03	2.428E+04
State 2	1.0%	1,100,000,000	25,000,000	-18,922,104	1,085,911,753	4.385E-03	-8.297E+04
State 3	0.5%	1,100,000,000	50,000,000	-43,922,104	1,067,298,274	4.813E-03	-2.114E+05
State 4	47.0%	1,000,000,000	0	6,077,896	1,004,525,231	6.587E-03	4.004E+04
State 5	2.0%	1,000,000,000	50,000,000	-43,922,104	967,298,274	7.935E-03	-3.485E+05
State 6	1.0%	1,000,000,000	100,000,000	-93,922,104	930,071,318	9.558E-03	-8.977E+05
State 7	23.5%	900,000,000	0	6,077,896	904,525,231	1.086E-02	6.601E+04
State 8	1.0%	900,000,000	100,000,000	-93,922,104	830,071,318	1.576E-02	-1.480E+06
State 9	0.5%	900,000,000	100,000,000	-93,922,104	830,071,318	1.576E-02	-1.480E+06
State 10							
Mean	100%	1,000,000,000	4,000,000	2,077,896	1,001,547,075		6.643E-04

Reinsurer C	Probability	W	L	$-L+p$	$W-\theta L+\theta p$	$u'(W-\theta L+\theta p)$	$u'(\,)(-L+p)$
State 1	87.0%	1,000,000,000	0	6,077,896	994,474,668	6.927E-03	4.210E+04
State 2	10.0%	1,000,000,000	50,000,000	-43,922,104	1,039,928,982	5.519E-03	-2.424E+05
State 3	3.0%	1,000,000,000	100,000,000	-93,922,104	1,085,383,296	4.397E-03	-4.129E+05
State 4							
State 5							
State 6							
State 7							
State 8							
State 9							
State 10							
Mean	100%	1,000,000,000	8,000,000	-1,922,104	1,001,747,358		-9.633E-03

APPENDIX A

The Imposture of Risk-Adjusted Discounting

One widely used and highly respected textbook on finance opens with the definitive and programmatic statement:

> To calculate present value, we discount expected payoffs by the rate of return offered by equivalent investment alternatives in the capital market. The rate of return is often referred to as the discount rate, hurdle rate, or opportunity cost of capital. [Brealey and Myers, 2002, p. 15]

That the payoffs (here assumed to be positive, or at least nonnegative) may be stochastic is unimportant. As long as one knows which investments are equivalent to the one in question, and at what rate of return these equivalent investments are discounted, one needs only to discount the "expected payoffs." On the same page the authors repeat this statement, explicitly mentioning risky investments:

> Here we can invoke a second basic financial principle: *A safe dollar is worth more than a risky one.* Most investors avoid risk when they can do so without sacrificing return. However, the concepts of present value and the opportunity cost of capital still make sense for risky investments. It is still proper to discount the payoff by the rate of return offered by an equivalent investment. But we have to think of *expected* payoffs and *expected* rates of return on other investments.

One can hardly disagree with the second basic principle; if the present values of the expected cash flows of a riskless investment and a risky investment are equal, the risky investment should be worth less. But the authors see no way to make the risky investment worth less other than to discount it at a greater rate of return, a return equal to that "offered by an equivalent investment." One should a least wonder how to tell which investments are equivalent.[21] And if there are equivalents, is there one that is

[21] Why do the authors even suggest that not all investments are equivalent when they intend to argue that the Capital Asset Pricing Model renders them all β-equivalent?

standard and priced by itself? If there is no standard, how do we avoid circular reasoning, viz., that A should be discounted at 15% because of B and that B should be discounted at 15% because of A? But even apart from the matter of circular reasoning, we can here show that discounting expected payoffs at risk-adjusted rates of return leads to inconsistencies. This will help us to see that conceiving present value as a random variable is necessary for consistently valuing stochastic cash flows.

Consider again the simplest stochastic cash flow of Section 3, the receipt of either 120 or 80 dollars one year from now, depending on a coin toss. The source of this example [D'Arcy, 1999, p. 23] assumed a risk-free discount rate of seven percent per year, and a risk-adjusted discount rate of twelve. The expected flow of 100 dollars would be discounted at twelve percent per year to yield a present value of 89.29 dollars. Though this seems reasonable, inconsistencies appear when one generalizes the problem to time t. The general present-value formula is $\$100/(1.12)^t$; but for $t > 4.89$ years this will be less than the present value of the tails-dependent cash flow $\$80/(1.07)^t$. Wishing to avoid this inconsistency, many would resort to making the risk-adjusted discount rate to vary with t, i.e., $\$100/(1+r(t))^t$ However, as t increases, the discount rate must decrease, approaching the risk-free lower bound of seven percent per year. And as t approaches zero, $\$100/(1+r(t))^t$ approaches \$100 (unless $r(t)$ approaches infinity), and the expected profit vanishes. Moreover, if the discount rate had to vary with time, equivalent investment alternatives would be fewer and harder to identify.

Finance textbooks assume that the cash flows to be discounted are nonnegative.

Normally this is true of investments, especially of stocks and bonds. But stochastic cash flows are more comprehensive. We could think of insured losses as negative payoffs, and the coin-toss flow would be the receipt of either –120 or –80 dollars one year from now. An insurer would like to "pay" less than the present value of –100 dollars, and this would require discounting at a rate less than the risk-free seven percent per year.[22] But as with the example with positive amounts, one would have to take care not to pierce the envelope, especially its lower bound of $-\$120/(1.07)^t$; and the problem of discounting too much for large t and too little for small t persists.

Inconsistency appears in yet another form. What is to prevent us from decomposing the \$120/\$80 cash flow into a degenerate cash flow of x dollars one year from now and a stochastic cash flow of either \$120–x or \$80–x one year from now, depending on the coin toss? It is hard to imagine why the "\$120–x or \$80–x" flow would not be equivalent to the same investments as those to which the "\$120 or \$80" flow is equivalent. Then it too would be discounted at twelve percent per year. But the x dollars are now discounted at seven percent per year, and the present value depends on x:

$$PV = x/(1.07)^t + (\$100 - x)/(1 + r(t))^t$$

Should a change in the level of the stochastic cash flow place the asset into a different class of equivalents, and justify a risk-adjusted rate that now depends not just on t, but also on x, i.e., $r(t,x)$? Especially troublesome is the case of x = \$100, the expected

[22] So according to Robert Bustic [1988, p. 149]: "The risk-adjusted interest rate [for loss reserves], being lower than yield rates available in the market, falls between the two extremes of not discounting (a zero interest rate) and discounting with a market rate." On p. 169 he estimates the risk-adjusted rate to be about four percent per year lower than the risk-free rate. That U.S. Treasury rates might become so low that one would discount loss reserves at a negative rate was unimaginable to him at the time.

payoff, in which case $PV = \$100/(1.07)^t$. This clearly violates the second principle of Brealey and Myers, for risky dollars are now worth as much as safe ones.

The only answer to this form of inconsistency is to forbid decomposing cash flows, or more accurately, to insist that each stochastic cash flow has one and only one suitable decomposition. This actually is the genesis of capital allocation. If stochastic cash flow C has negative outcomes, one allocates enough capital k to offset its negative outcomes, or at least to render them insignificant. Then one discounts the combination of the degenerate cash flow $-k$ and (the expected payoff of) the stochastic cash flow $k+C$.[23] But which degenerate cash flow k is the suitable one? In particular, if C has no lower bound, how much capital is required to render the remaining negative outcomes insignificant? Only to true believers are these inconsistencies trivial. The rest, not knowing a better way, may acquiesce to risk-adjusted discounting and capital allocation; but they suspect that such practices are arbitrary.

On two principles the foregoing argument has been based. The first principle is that the value of a stochastic cash flow must lie within the minimum and maximum values of its outcome-dependent cash flows. In other words, if random variable X is the present value of a stochastic cash flow, as defined in Section 3, and if for some real number a $P(X < a) = 0$, then the value of the stochastic cash flow must be greater than or equal to a. Similarly, if for some b $P(X > b) = 0$, the value must be less than or equal to b. In the

[23] Some practitioners discount the expected flows separately, C at the risk-free rate and the allocated capital at some ROE rate. This only underscores the absence of theory from risk-adjusted discounting.

extreme, if random variable X is degenerate, i.e., for some a $P(X = a) = 1$, the value of the stochastic cash flow must be a. For example, if all ten outcomes of Exhibit 1 lay on the $100 isobar, the cash flow would have to be worth 100 dollars.

Some disagree with this inference, arguing that the several points extending over the isobar create a timing risk that no single point has. However, the risk that affects value resides not in the timing of the outcomes, but rather in their present value. And the present value is always 100 dollars. Others argue that if the present-value operator, or the coordinate system, changes, the ten outcomes will lie on different isobars. Though this is true, it ignores what was said in Section 2, viz., that the present-value operator is momentary and does not second-guess itself. Moreover, if this argument were valid, it would prove too much. For it would work even if the cash flow consisted of just one point on the $100 isobar, for instance, the point (5, $117.69). One might just as well argue that due to "interest-rate risk" this one point ought to be worth somewhat less than 100 dollars.

Donald Mango has coined a phrase that poignantly addresses this issue:

> ... involves discounting cash flows at a default-free rate. Scenario analysis is built upon the premise that possible, realizable, plausible outcomes can be generated and analyzed. For the entire process to work, each generated scenario is "**conditionally certain**": given the scenario occurs, its outcome is certain. Where it is not, the entire practice of simulation modeling would be undermined by "meta-uncertainty."
>
> Uncertainty for the contract in total is represented in the distribution across all modeled scenarios, and the probability weights assigned to those scenarios. In other words, *uncertainty is reflected between the scenarios, not within them.* Given conditional certainty, scenario cash flows can be discounted at a default-free rate. [2003, p. 355f.]

Conditional certainty is a striking concept. For our basic space S, consisting of a countable number of elementary outcomes ξ_i, the Theorem of Total Probability states:

$$P(A)=\sum_i P(A|\xi_i)P(\xi_i)$$

We can analogize from the probability of event A to the value of stochastic cash flow C, using present value as the conditional operator:

$$V(C)=\sum_i PV(C|\xi_i)P(\xi_i)$$

But this would be the expectation of the present value. One must adjust the probabilities $P(\xi_i)$ to obtain a risk-adjusted value for C, a value that will lie between any minimum and maximum since it is a weighted average.[24]

The second principle is that to change the level of a stochastic cash flow is to change its value by the amount of the level. In symbols, $V(k+C) = k+V(C)$. Philosophically this means that no point on the real continuum is special, not even zero.[25] Negative one million dollars is merely one million dollars less than zero, and two million dollars less than one million. It may mean bankruptcy to an accountant,[26] but for valuation it is just another real number. The principle means also that we can decompose cash flows. In fact, we could generalize this principle into linearity.[27] For stochastic cash flows C and D, and constants α and β, $V[\alpha C+\beta D]=\alpha V[C]+\beta V[D]$. A non-linear valuation operator enables arbitrage; nevertheless, finance textbooks are devoted more to risk-adjusted

[24] Adjusted probabilities reappear in the discussion of Section 5 concerning state prices.
[25] As mentioned in Section 2, this renders valuation independent of solvency.
[26] Some argue that negative wealth is not possible with bankruptcy laws. However, these are positive laws, and something *de jure humano* is not worthy of the name "theory." Moreover, bankruptcy law does not make everyone's wealth the greater of zero and book value. One whose book value is negative must file for bankruptcy, turn his assets over to a court for liquidation to his creditors, and possibly suffer long-term consequences (e.g., ineligibility for credit, imprisonment). Furthermore, creditors and society at large bear the debt that remains after liquidation. So bankruptcy laws recognize the reality of negative wealth.
[27] Paradoxically, linearity obtains only in an equilibrium (Section 5). Though the generalization is important, for the purpose of this appendix it is unnecessary.

discounting than to the avoidance of arbitrage. Risk-adjusted discounting violates these two principles, and additionally, cannot build realistic risk loads into short-duration risks.

Finally, we comment about the notions of rate of return and cost of capital. For centuries individuals and banks have lent money "for so many years at such an interest rate per year, payable so many times per year." So everyone became accustomed to measuring the cost of money in percent per year, and to imagining that borrowed money worked at that rate of return. In Section 2 we argued for the present-value function $v(t)$. Rather than say, for example, that the cost of dollars is five percent per year, we should say that the cost of one dollar t years from now is $v(t)$ dollars today. The cost of money later is not percent per year, but rather an amount of money now. But even if one insists on cost of capital as percent per year, one cannot apply it to stochastic cash flows. Money loaned out at five percent per year is money that is eventually repaid. But capital allocated to a risky project is capital that stands ready to be sacrificed. Even if it were proper to say, "My money works at five percent per year while it's away from me," it would not necessarily be proper to say, "My money works at fifteen percent per year while it's risking its life."[28]

[28] For similar opinions that a stochastic cash flow cannot be valued by adjusting the discount rate see Halliwell [2001, §3], Schnapp [2001], and Van Slyke [1995 and 1999]. Schnapp [2001, §8] writes, " ... the certainty equivalent price for future outcomes can be obtained by discounting the uncertain outcomes to present value and then determining the certainty equivalent price." So too Van Slyke: "The cost of uncertainty in the future is treated as a real cost, and is discounted at only the time value of money indicated by the currency markets." [1995, p. 617] and " ... a distinction between the time value of money and the cost of risk. The time value of money is recognized by replacing all outcomes that may be realized at future times with their equivalent values in current dollars." [1999, p. 140]

APPENDIX B

Properties of the Expected Utility Function

Here follows a proof of the claim in Section 5 that if the net cash flow, i.e., the cash flow minus its price, or $X - q$, has both upside and downside potential, then the graph of the equation $f(\theta) = E[u(W + \theta X - \theta q)]$ looks like a concave downward parabola (provided that u is a risk-averse utility function). Hence, $f(\theta)$ has one and only one maximum. This holds true regardless of the joint distribution of W and X.

The second derivative is $f''(\theta) = E[u''(W + \theta X - \theta q)(X - q)^2]$. As a risk-averse utility function, u'' is negative. Therefore, unless $\text{Prob}[X = q] = 1$ (contrary to the downside-upside assumption), f'' is everywhere negative. This establishes that f is concave downward, or equivalently, that f' strictly decreases. Hence, there can be at most one critical value, a value θ at which $f'(\theta) = 0$, at which value f is a maximum. But not established is that there must be at least one critical point. It will be proven that $\lim_{\theta \to \pm\infty} f'(\theta) = \mu\infty$. The continuity of f' will then guarantee a zero.

For the proof that $\lim_{\theta \to +\infty} f'(\theta) = -\infty$, one may assume that θ is positive. Since the stochastic cash flow has downside potential at price q, there exists some $q_0 < q$ such that $F_X(q_0) = \text{Prob}[X \le q_0] > 0$. And since it has upside potential, $1 - F_X(q) = \text{Prob}[X > q] > 0$. And the probability of the middle interval, $q_0 < X \le q$, is

greater than or equal to zero. So the first step is to express $f'(\theta)$ in terms of conditional probability:

$$\begin{aligned} f'(\theta) &= E[u'(W+\theta X-\theta q)(X-q)] \\ &= E[u'(W+\theta X-\theta q)(X-q)\,|\,X \le q_0]F_X(q_0) \\ &+ E[u'(W+\theta X-\theta q)(X-q)\,|\,q_0 < X \le q](F_X(q)-F_X(q_0)) \\ &+ E[u'(W+\theta X-\theta q)(X-q)\,|\,q < X](1-F_X(q)) \end{aligned}$$

In the middle interval, $q_0 < X \le q$, $u'(W+\theta X-\theta q)(X-q) \le 0$. For u' is everywhere positive, and in that interval $X-q$ is less than or equal to zero; hence, the product is less than or equal to zero. And in the upper interval, $X-q>0$. Since θ is positive, $\theta X-\theta q>0$ and $W+\theta X-\theta q>W$. u' is positive, but strictly decreasing. Hence, $u'(W+\theta X-\theta q)<u'(W)$, and $u'(W+\theta X-\theta q)(X-q)<u'(W)(X-q)$. Therefore:

$$\begin{aligned} f'(\theta) &= E[u'(W+\theta X-\theta q)(X-q)\,|\,X \le q_0]F_X(q_0) \\ &+ E[u'(W+\theta X-\theta q)(X-q)\,|\,q_0 < X \le q](F_X(q)-F_X(q_0)) \\ &+ E[u'(W+\theta X-\theta q)(X-q)\,|\,q < X](1-F_X(q)) \\ &\le E[u'(W+\theta X-\theta q)(X-q)\,|\,X \le q_0]F_X(q_0) \\ &+ E[u'(W+\theta X-\theta q)(X-q)\,|\,q < X](1-F_X(q)) \\ &< E[u'(W+\theta X-\theta q)(X-q)\,|\,X \le q_0]F_X(q_0) \\ &+ E[u'(W)(X-q)\,|\,q < X](1-F_X(q)) \end{aligned}$$

The next step concerns the lower interval. Since here $X \le q_0$, $X-q \le q_0-q$. u' is positive, so the following are true:

$$\begin{aligned} u'(W+\theta X-\theta q)(X-q) &\le u'(W+\theta X-\theta q)(q_0-q) \\ E[u'(W+\theta X-\theta q)(X-q)\,|\,X \le q_0] &\le E[u'(W+\theta X-\theta q)\,|\,X \le q_0](q_0-q) \\ &= E[-u'(W+\theta X-\theta q)\,|\,X \le q_0](q-q_0) \end{aligned}$$

However, $-u'$ is an increasing, but concave downward function. It meets the conditions of Jensen's inequality (Section 4):

$$\begin{aligned} E[u'(W+\theta X-\theta q)(X-q)|X\le q_0]&\le E[-u'(W+\theta X-\theta q)|X\le q_0](q-q_0)\\ &\le -u'(E[(W+\theta X-\theta q)|X\le q_0])(q-q_0)\\ &= -u'(E[W|X\le q_0]+\theta E[X-q|X\le q_0])(q-q_0) \end{aligned}$$

So finally:

$$\begin{aligned} f'(\theta) &< E[u'(W+\theta X-\theta q)(X-q)|X\le q_0]F_X(q_0)\\ &\quad + E[u'(W)(X-q)|q<X](1-F_X(q))\\ &< -u'(E[W|X\le q_0]+\theta E[X-q|X\le q_0])(q-q_0)F_X(q_0)\\ &\quad + E[u'(W)(X-q)|q<X](1-F_X(q)) \end{aligned}$$

θ appears only once on the right side of the inequality. Its coefficient, $E[X-q|X\le q_0]$, is negative. So as θ approaches ∞, the argument of $-u'$ approaches $-\infty$. Since $-u'$ is an increasing, concave downward function, it must approach $-\infty$. And its coefficient, $(q-q_0)F_X(q_0)$, is positive. Therefore, $\lim_{\theta\to+\infty} f'(\theta)=-\infty$.

Duality makes the proof of the opposite limit simple. Since X has both upside and downside potential at price q, $-X$ has both upside and downside potential at price $-q$. What is a limit as $\theta\to-\infty$ with cash flow X at price q can be expressed as a limit as $-\theta=\zeta\to+\infty$ with cash flow $-X$ at price $-q$:

$$\begin{aligned} \lim_{\theta\to-\infty} f'(\theta) &= \lim_{\theta\to-\infty} E[u'(W+\theta X-\theta q)(X-q)]\\ &= \lim_{\zeta\to+\infty} E[u'(W+(-\zeta)X-(-\zeta)q)(X-q)]\\ &= \lim_{\zeta\to+\infty} E[u'(W+\zeta(-X)-\zeta(-q))((-X)-(-q))]\cdot(-1)\\ &= -\infty\cdot(-1)\\ &= +\infty \end{aligned}$$

Since f' is continuous and strictly decreasing, but ranges over all the real numbers, it must have one and only one zero. So f has one and only one maximum.

APPENDIX C

Exponential Utility

Is there an ideal form of utility? Section 4 stated commonly accepted properties, viz., that $u(w)$ should be increasing, twice differentiable, and concave downward. One could argue that quadratic, power, and logarithmic functions have these properties. However, the quadratic function has a maximum utility at its vertex; utilities of greater wealth are undefined. And power and logarithmic functions are undefined for negative wealth. Arguments in Appendix A to the contrary, many appeal to bankruptcy law for putting a floor of zero under the wealth of an economic agent. But even here power and logarithmic functions fail, because utility distributions could have probability masses at $u(0) = -\infty$. The ideal utility function would be defined for all real numbers, and its first derivative would approach zero as wealth approached positive infinity. It is hard to devise any function with all these properties other than the exponential.

Hans Gerber [1979, p. 70] proposes five desirable properties for premium calculation. One of these properties is additivity, viz., that the premium for the combination of two independent risks should equal the sum of the premiums for the single risks. He demonstrates that "the principle of zero utility" has this property if and only if the utility function is linear or exponential. Gerber's principle is equivalent to our formulation of utility in Section 4. However, in Section 5 we recommended a principle of maximal utility, according to which an economic agent whose stochastic wealth W is in

equilibrium should value stochastic cash flow X according to the formula:

$$q_X = \frac{E[u'(W)X]}{E[u'(W)]}$$

Being linear in X, this formula is additive for any utility function. Moreover, the formula holds even when the risks are not independent, which exposes the weakness that Gerber's principle is not in general additive.

In the case of exponential utility the formula above becomes:

$$q_X = \frac{E[-e^{-aW}X]}{E[-e^{-aW}]} = \frac{E[e^{-aW}X]}{E[e^{-aW}]}$$

Hans Bühlmann [1980, p. 58] notes that when X and $W-X$ are independent:

$$q_X = \frac{E[e^{-aW}X]}{E[e^{-aW}]} = \frac{E[e^{-aX}X]}{E[e^{-aX}]}$$

The right side of the last equation is known as the Esscher transform of X. This form is especially attractive to insurance, since it allows "quantum" cash flows, i.e., cash flows independent of the rest of the portfolio, to be valued on their own. This suggests that the ideal utility function should allow quantum cash flows to be valued on their own; otherwise, one would have to know potentially everything in order to value anything. Unfortunately, we have not been able to prove that only exponential utility allows for this.

However, there is a related property. If X is independent of $W-X$, then for any constant k it is also independent of $k+W-X$. The principle is that the level of wealth should be irrelevant to equilibrium; in other words, if W is an equilibrium wealth, then so too should

be $k+W$. This comports with the second principle of Appendix A, viz., that no point on the real continuum is special, not even zero. This will offend many "solvency-minded" persons, but even in everyday matters value is unrelated to one's financial condition. Asking a valuation formula to depend on k is like asking a shopkeeper to charge lower prices to the poor than to the rich. Therefore, we recommend that the valuations of an ideal utility function should be invariant to level; in symbols, for all k, W, and X:

$$\frac{E[u'(k+W)X]}{E[u'(k+W)\ ]} = \frac{E[u'(W)X]}{E[u'(W)\ ]}$$

As we are about to see, this recommendation makes demands on the form of u.

To begin, this equation cannot be true for all X unless:

$$\frac{u'(k+W)}{E[u'(k+W)]} = \frac{u'(W)}{E[u'(W)]}$$

And one can substitute tW for W, for arbitrary real number t. Hence, u must ensure for all k, t, and W the equality:

$$\frac{u'(k+tW)}{E[u'(k+tW)]} = \frac{u'(tW)}{E[u'(tW)]}$$

Now differentiate the equation with respect to t:

$$\frac{E[u'(k+tW)]u''(k+tW)W - E[u''(k+tW)W]u'(k+tW)}{E[u'(k+tW)]^2} = \frac{E[u'(tW)]u''(tW)W - E[u''(tW)W]u'(tW)}{E[u'(tW)]^2}$$

In particular, this equation is true for $t = 0$:

$$\frac{E[u'(k)]u''(k)W - E[u''(k)W]u'(k)}{E[u'(k)]^2} = \frac{E[u'(0)]u''(0)W - E[u''(0)W]u'(0)}{E[u'(0)]^2}$$

And its simplified form is:

$$\frac{u''(k)(W - E[W])}{u'(k)} = \frac{u''(0)(W - E[W])}{u'(0)}$$

But this can not hold for all W unless for all k:

$$\frac{u''(k)}{u'(k)} = \frac{u''(0)}{u'(0)}$$

Only linear and exponential functions solve this equation, which expresses constant absolute risk aversion (Longley-Cook [1998, p. 90]). And by standardizing $u(0)$ and $u'(0)$ (Gerber [1979, p. 68] and Halliwell [1999, §6]) the linear function becomes the limit as a approaches zero of the exponential function $u(w) = \left(1 - e^{-aw}\right)/a$.

The ideal form of utility is (linear-)exponential. It is hard to devise other functions that possess the commonly accepted properties. Exponential utility allows for independent risks to be valued on their own, and perhaps it alone has this distinction. Finally, it alone renders valuation independent of the arbitrary level of an agent's stochastic wealth.

APPENDIX D

Cumulants and the Esscher Transform

In accordance with Appendix C, the formula for the value of a quantum cash flow (a cash flow that is independent from the rest of an agent's portfolio) can be transformed:

$$\begin{aligned} q_X &= \frac{E\left[Xe^{-aX}\right]}{E\left[e^{-aX}\right]} \\ &= -\frac{d}{da}\ln E\left[e^{-aX}\right] \\ &= -\frac{d}{da}\ln M_X(-a) \\ &= -\frac{d}{da}\psi_X(-a) \end{aligned}$$

The function $\psi_X(a)$, the logarithm of the moment generating function, is called the cumulant generating function (Daykin [1994, p. 23] and Halliwell [1999, Appendix C]), not to be confused with the state-price random variable Ψ of Section 5. Its derivatives evaluated at zero are called the cumulants of X; they are the κ_i of the Maclaurin-series expansion $\psi_X(t)=\sum_{i=1}^{\infty}\frac{\kappa_i t^i}{i!}$. The first three cumulants are the mean, the variance, and the skewness; the fourth cumulant is the kurtosis with an adjustment:

$$\begin{aligned} \kappa_1 &= E[X]=\mu \\ \kappa_2 &= E\left[(X-\mu)^2\right]=\sigma^2 \\ \kappa_3 &= E\left[(X-\mu)^3\right] \\ \kappa_4 &= E\left[(X-\mu)^4\right]-3\sigma^4 \end{aligned}$$

All the cumulants of the normal distribution beyond the second are zero. The cumulants of a sum of independent random variables equal the sums of the variables' cumulants.

So the value of a quantum cash flow can be expressed in terms of its cumulants:

$$\begin{aligned} q_X &= -\frac{d}{da}\psi_X(-a) \\ &= -\frac{d}{da}\sum_{i=1}^{\infty}(-1)^i \frac{\kappa_i a^i}{i!} \\ &= \sum_{i=1}^{\infty}(-1)^{i-1} \frac{\kappa_i a^{i-1}}{(i-1)!} \\ &= \kappa_1 - \kappa_2 a + \frac{\kappa_3}{2}a^2 - \frac{\kappa_4}{6}a^3 + \mathrm{K} \\ &= \mu - \sigma^2 a + \frac{\kappa_3}{2}a^2 - \frac{\kappa_4}{6}a^3 + \mathrm{K} \end{aligned}$$

From this follows the formula for the expected profit:

$$\begin{aligned} \pi_X &= E[X] - q_X \\ &= \mu - q_X \\ &= \sigma^2 a - \frac{\kappa_3}{2}a^2 + \frac{\kappa_4}{6}a^3 - \mathrm{K} \end{aligned}$$

As a first-order approximation, $\pi_X \approx \sigma^2 a$, which holds exactly for normally distributed X.

One can generalize the cumulant generating function and its Maclaurin series to two random variables:

$$\begin{aligned} \psi_{X,W}(s,t) &= \ln E\left[e^{sX+tW}\right] \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\kappa_{ij}}{i!\,j!}s^i t^j \end{aligned}$$

The derivative of this function with respect to s, evaluated at $(0, -a)$, is the general formula for value:

$$\frac{\partial}{\partial s}\psi_{X,W}(s,t) = \frac{E\left[Xe^{sX+tW}\right]}{E\left[e^{sX+tW}\right]} = \sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\frac{\kappa_{ij}}{(i-1)!\,j!}s^{i-1}t^{j}$$

$$\frac{\partial}{\partial s}\psi_{X,W}(s,t)\Big|_{\substack{s=0\\t=-a}} = \frac{E\left[Xe^{-aW}\right]}{E\left[e^{-aW}\right]} = \sum_{j=0}^{\infty}\frac{\kappa_{1j}}{j!}(-a)^{j}$$

Somewhat tediously, the formulae for the following cumulants can be derived:[29]

$$\kappa_{10} = E[X] = \mu_X$$
$$\kappa_{11} = E[(X-\mu_X)(W-\mu_W)] = Cov[X,W] = \sigma_{X,W}$$
$$\kappa_{12} = E\left[(X-\mu_X)(W-\mu_W)^2\right]$$
$$\kappa_{13} = E\left[(X-\mu_X)(W-\mu_W)^3\right] - 3\sigma_{X,W}\sigma_W^2$$

By replacing X with W one obtains the cumulants of W in keeping with the earlier formulae. The formula for expected profit is:

$$\begin{aligned}\pi_X &= E[X] - q_X \\ &= E[X] - \left(\kappa_{10} - \kappa_{11}a + \frac{\kappa_{12}}{2}a^2 - \frac{\kappa_{13}}{6}a^3 + \mathrm{K}\right) \\ &= E[X] - \left(E[X] - Cov[X,W]a + \frac{\kappa_{12}}{2}a^2 - \frac{\kappa_{13}}{6}a^3 + \mathrm{K}\right) \\ &= Cov[X,W]a - \frac{\kappa_{12}}{2}a^2 + \frac{\kappa_{13}}{6}a^3 - \mathrm{K}\end{aligned}$$

Though beautiful, the formula is often impractical, not just because high-order cumulants are unfamiliar, but mainly because it may converge slowly. In the case of Reinsurer A of Exhibit 10, q_X is nowhere near the total premium of 5.1 million dollars after four terms,

[29] Kozik and Larson [2001, 58-63] derive similar formulae, but in terms of rates of return. Though much can be learned from their derivation, rate of return is an accounting notion that introduces an arbitrary time horizon (viz., one year). Our present-value perspective (Sections 2 and 3), which keeps to monetary units (dollars) and impounds all the future into the present moment, avoids the risk-adjusted discounting implicit in the rate-of-return approach (Appendix A).

i.e., after the κ_{13} term. Since the order of the magnitude of the j^{th} term of the series for q_X is $\kappa_{1j}a^j \approx W^j a^j \approx (1{,}000{,}000{,}000 \times 5.000\text{E}-09)^j = 5^j$, the factorial in the denominator does not begin to dampen the series before the fifth term.

Nevertheless, the formula provides the general first-order approximation, $\pi_X \approx \sigma_{X,W} a$, which is exact when X and W are normally distributed. That expected profit, or risk load, is approximately proportional to covariance is true to the spirit of the Capital Asset Pricing Model. Since we can express the approximation with the correlation coefficient as $\pi_X \approx \rho_{X,W}\sigma_X\sigma_W a$, one may say that risk load is proportional to standard deviation (Kreps [1990, p. 198]). And if X is independent of W, the risk load should be zero. However, one must distinguish the independence of X from W and the independence of X from W–X. In the latter case,

$$\begin{aligned} Cov[X,W] &= Cov[X, X+W-X] \\ &= Cov[X,X] + Cov[X, W-X] \\ &= Var[X] \end{aligned}$$

Then the risk load is proportional to variance. From the premise that X should be small relative to W some[30] infer that the covariance and the risk load should be effectively zero. The premise itself is debatable; however, the proper conclusion is merely that π_X should be small relative to π_W. It certainly is not negligible relative to σ_X^2

[30] E.g., Wang [2001, §2]: "For an insurance market in which any individual risk is negligible relative to the aggregate risk, under the assumption that X and Z–X are independent, the equilibrium premium for X equals the expected loss without risk loading."

The Aggregation and Correlation of Reinsurance Exposure

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

The Aggregation and Correlation of Reinsurance 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 a reinsurer. We include most of the standard exposures as well as property catastrophe exposure. Next we show how this aggregate loss distribution can be used to determine the needed capital, and its cost, for a reinsurer. Finally we show how to calculate the capacity charges for individual reinsurance contracts that will allow the reinsurer to recover its cost of capital. We demonstrate the use of this methodology on some illustrative reinsurance contracts. We believe this methodology can be used in practice by most reinsurers.

Authors

Glenn Meyers heads the Special Markets Products Division at Insurance Services Office. Since joining ISO in 1988 he has led the technical development of ISO's increased limits ratemaking, PSOLD – ISO's commercial property size of loss database, MILD – ISO's reinsurance exposure rating product, ISO's joint loss and ALAE distributions and ISO's Underwriting Risk Model. Glenn's academic background includes a B.S. in Mathematics and Physics from Alma College, a M.A. in Mathematics from Oakland University and a Ph.D. in Mathematics from SUNY at Albany. He is a FCAS and MAAA. He is currently serving the CAS on the Dynamic Risk Modeling Committee and the Future Education Task Force.

Fred Klinker works with Glenn in the Special Markets Products Division at Insurance Services Office. Prior to coming to ISO, he worked for seventeen years in Actuarial Research at CIGNA P&C, doing financial, mathematical, and statistical modeling. He has a BS/BA in Physics/Math from the University of Rochester and graduate work in Physics and Math from the Universities of Rochester and Pennsylvania. He is a FCAS and MAAA.

David Lalonde is Senior Vice President at AIR Worldwide Corporation responsible for the Services and Analytics group, providing Catastrophe Loss Analysis Services (CLAS™) and Risk Transfer Services (RTS™). He has been involved in the development and use of catastrophe modeling technology since 1995 and regularly assists clients in use of models for exposure management, ratemaking and underwriting, risk transfer and the securitization of insurance risk, portfolio optimization, and dynamic risk evaluation and analysis models. Mr. Lalonde is a FCAS, FCIA, MAAA. Mr. Lalonde earned his B.Math. (Honours) in Actuarial Science with Statistics from the University of Waterloo. . He is currently serving on the CAS Advisory Committee on Securitization/Risk Financing.

1. Introduction

This paper has three objectives:

1. Demonstrate a practical method to determine the distribution of a reinsurer's aggregate loss payments. This includes not only losses from the contracts it currently is reinsuring, 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 reinsured.
2. Using the results of Objective #1, demonstrate how to determine the amount of capital needed for a reinsurance company based on its risk of loss.
3. Using the results of Objective #2 demonstrate how to determine the capacity charge for a new reinsurance contract.

We will illustrate the use of our model and methodology on a portfolio of reinsurance contracts. The parameters for the loss models were obtained from analyses by Insurance Services Office (ISO) and AIR Worldwide Corporation (AIR).

The exposures for these contracts were obtained from the annual statements for several primary insurers and from data reported to ISO. Using the descriptions of the reinsurance programs that were reported to A.M. Best Company, we modified the loss models accordingly.

We treat the time value of money by assuming a fixed risk-free interest rate. While the assets of a reinsurer are not always risk-free, a full treatment of asset risk is beyond the scope of this paper. Thus, we should expect reinsurers 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 a reinsurer's distribution of underwriting losses. This description will include ways to model dependencies among the various reinsurance contracts. 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 calculating a capacity charge. 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 the capacity charge 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 Reinsurer 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 reinsured loss in terms of the underlying claim count and claim severity distributions for each reinsured contract. We describe this model by the following simulation algorithm.

Simulation Algorithm #1

Step

1. For each reinsurance contract, 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 contract 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 contract h.
3. Set $X = \sum_h X_h$ = Loss for the reinsurer.

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 reinsurance contract 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 contract 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 contract 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 reinsurer.

The extra step of multiplying all the losses by a random β adds variability in a way that losses for each reinsurance contract 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 reinsurance contracts. 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.

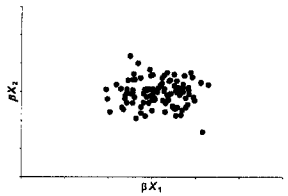


Figure 2

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.

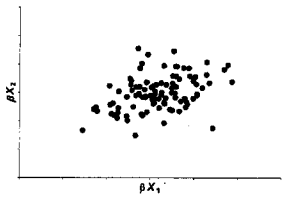


Figure 3

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.

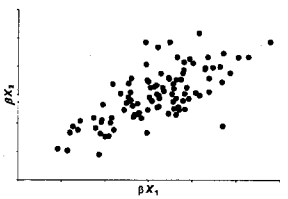
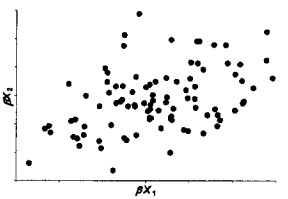


Figure 4

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 a reinsurer. Here is a summary of the main features of this model.

- It is necessary to hold capital for uncertain losses in expired reinsurance contracts. Thus the model treats unpaid losses from both new and expired reinsurance contracts from prior accident years

- We use separate parameter uncertainty multipliers for both claim frequency and claim severity. For reinsurance contract 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 contract h.

- Each reinsurance contract is assigned to a distinct "covariance group" according to the line of business that it covers. (Granted, some reinsurance contracts cover multiple lines, but in this paper, we use a narrower definition of "contract.") 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 business in the same covariance group, but they increase and decrease together.

- The random claim severity multiplier, β, is applied uniformly across all contracts.

- 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 contracts 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, reinsurance contract 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 reinsurance contract h and accident year y in covariance group i.
   - 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 reinsurer.

We now describe our parameterization of this model.

- For the non-catastrophe reinsurance contracts, 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 reinsurance contracts, we removed those claims that are already settled.

- Reinsurers often write multiple contracts, covering different layers, with a single insurer. For example, one reinsurance contract will cover 50% of a lower layer, and another contract will cover 80% of a higher layer. We treat such arrangements as a single contract and adjust the claim severity distribution accordingly.

- We use the negative binomial distribution to model claim counts. The expected claim count will depend on the reinsurer'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 a reinsurer'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 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 or reinsurer 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.

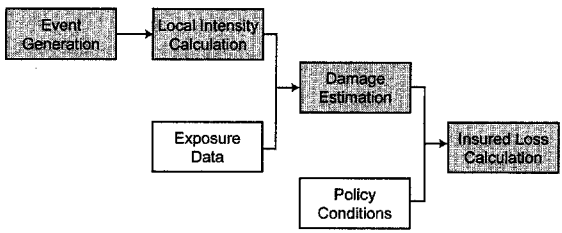
Fortunately, such events are infrequent. But it is exactly their infrequency that makes the estimation of losses from future catastrophes so difficult. The scarcity of historical loss data makes standard actuarial techniques of loss estimation inappropriate for quantifying catastrophe losses. Furthermore, 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 are 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 influence the behavior 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 reinsurance terms.

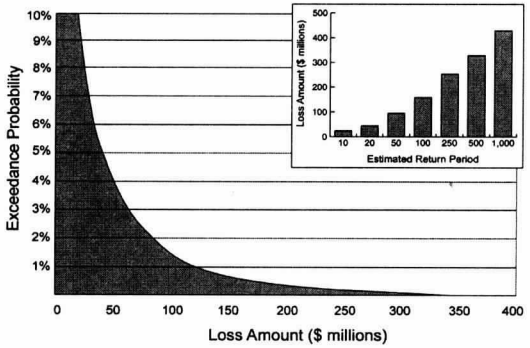
2.2.e Model Output

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 business and, within line of business, 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 reinsurance 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 amongst 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 a reinsurer depends solely on its aggregate loss distribution.

A reinsurer is exposed to underwriting risk not only from future claims on new business, but also from unsettled claims on past business. One must consider the underwriting risk from both sources when calculating the required capital. Larger claims tend to take longer to settle, and the underwriting loss model should reflect this.

Let X be the random variable for the reinsurer's total loss. Denote by $\rho(X)$ the total assets that the reinsurer needs to support its business[1]. Now some of the reinsurer'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 reinsurer 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 $\lambda \geq 0$ 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.

4. Calculating the Capacity Charge

As noted in the last section, a reinsurer needs to get capital from investors in order to attract business. The investors expect to be compensated in return for providing this capital at an expected rate of return that is somewhat higher than they would obtain for not exposing their capital to reinsurance risk. This additional return must come from the sum of the premiums charged to each individual reinsurance contract. The portion of this additional return for an individual reinsurance contract is called the capacity charge. In this section we give our formula for calculating the capacity charge.

Our formula is based on the underwriting strategy of establishing a target return on the additional capital needed to write this contract. We view the capacity charge as input into the underwriting decision. If the market will not allow the reinsurer to obtain this target return, the reinsurer should consider not writing the proposed contract.

We divide this section into two subsections. The first subsection gives our rationale for using this formula in terms of our chosen underwriting strategy. The second subsection gives our capacity charge formula.

4.1 A General Discussion of Capacity Charge Formulas

We take it as a given that a sound method of calculating capacity charges should lead to decisions that benefit the entire operation of a reinsurer.

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

Adding a reinsurance contract to a reinsurer's portfolio will increase the reinsurer's expected return on capital if and only if the contract's expected return on marginal capital (i.e., the contract's capacity charge divided by the additional capital needed to write the contract) is greater than the reinsurer's current expected return on capital.

This proposition provides a minimum standard on the capacity charge for *new* contracts. It says nothing about the capacity charge on existing contracts.

Proposition 2

Let the reinsurer'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 reinsurance contract is less than or equal to the reinsurer'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 contracts will have an expected return on marginal capital that is greater than the reinsurer'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 a reinsurance contract i, the random losses, X_i, for the contract are equal to a random number, U_i, times the exposure measure, e_i, for all possible values of e_i. Then, following Mildenhall [2002], the distribution of X_i is said to be ***homogeneous*** with respect to the exposure measure, e_i.

Proposition 3

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

Let the random loss, X_i, for the i^{th} reinsurance contract be a homogeneous random variable for each contract with respect to some exposure measure, e_i.

Let $X = \sum_i X_i$. Let the reinsurer'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 reinsurance contract is equal to the reinsurer'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 contracts.

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 reinsurer's overall return on capital.

Definition 2

The ***heterogeneity multiplier, HM,*** for a reinsurer is its needed capital divided by the sum of the marginal capitals for each contract in its reinsurance portfolio.

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

Our target capacity charge will be determined by a target return on marginal capital times the reinsurer'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 contracts to if the reinsurer is to make the most efficient use of its capital.
2. Propositions 2 and 3 – The sum of the marginal capitals over all reinsurance contracts is less than or equal to the total capital. The conditions that will force equality are not satisfied.

Note that the rationale underlying this charge depends on the individual reinsurance contracts being a small part of a reinsurer's portfolio, so that the smoothness criterion of Proposition 3 and 4 is a reasonable assumption.

4.2 The Capacity Charge Formula

If the underwriting result of all reinsurance contracts could be known within a year, we expect to release the capital at the end of the year, earning investment income on the capital. We would calculate the capacity charge as follows.

1. Establish a reference portfolio, Π, of existing contracts. Ideally this portfolio is updated as each new contract is accepted. But the year-end rush to book January 1 renewals makes this difficult to do in practice, so the portfolio will be set up according to a business plan.

2. Calculate the marginal capital for each contract in the reference portfolio. This is done by first calculating the capital needed for the reference portfolio according to Equation 1. Next we calculate the capital needed when a given contract is removed from the portfolio. The marginal capital for the given contract is the difference between the two capital calculations.

3. Calculate the heterogeneity multiplier, *HM*, by the formula:

$$HM = \frac{\text{Total capital for } \Pi}{\text{Sum of the marginal capitals of each contract in } \Pi}. \qquad (2)$$

4. For a prospective reinsurance contract, calculate the marginal capital, ΔC, needed when the contract is added to the reference portfolio.

5. Let r be the rate of return needed to attract the needed capital. Let i be the rate of return on invested assets. We expect $r > i$. The capacity charge, ΔP, for the prospective reinsurance contract is given by[2]:

$$\Delta P = \frac{(r-i)\cdot HM \cdot \Delta C}{(1+r)}. \qquad (3)$$

   Because of the way we defined the *HM*, the sum of the capacity charges will yield the reinsurer's desired return on its capital.

[2] This formula is special case of Equation 5, which is derived below.

As discussed above, the underwriting result of some reinsurance contracts can be uncertain for a period of several years. In this case the reinsurer must hold capital over this period to support these potential liabilities. This affects the calculation of the capacity charge in the following ways.

- The reference portfolio must contain the contracts that have expired but still have uncertain losses. The required capital for the reference portfolio must reflect the uncertainty in the ultimate losses from these unexpired contracts.
- When calculating the capacity charge, the reinsurer has to consider the fact that it has to hold additional capital in future years to support the contracts it is writing now. The cost of holding this capital over this extended period of time must be included in the capacity charge.

With these considerations in mind, we calculate the capacity charge as follows.

1. Establish a reference portfolio of existing contracts for this year, and for as many years in the future that the reinsurer expects uncertainty in its ultimate losses for contract written in this and prior years. Denote the portfolio for the current year by Π_0, the portfolio for next year by Π_1, and so on. These reference portfolios will contain current and expired and planned future reinsurance contracts.
2. Calculate the marginal capital for each reinsurance contract (current and expired) in each of portfolios, Π_0, Π_1....
3. For each portfolio, Π_n, calculate the heterogeneity multiplier, HM_n, by the formula.

$$HM_n = \frac{\text{Total capital for } \Pi_n}{\text{Sum of the marginal capitals of each contract in } \Pi_n}. \qquad (4)$$

4. For a prospective reinsurance contract, calculate the marginal capital, ΔC_n, needed when the contract is added to the n^{th} reference portfolio. Note that the contract is considered to be new in Π_0, but expired in Π_n, for $n > 0$.

5. Let r be the rate of return needed to attract the needed capital. Let i be the rate of return on invested assets. The capacity charge, ΔP, for the prospective reinsurance contract is given by:

$$\Delta P = \sum_{n=0}^{\infty} \frac{(r-i) \cdot HM_n \cdot \Delta C_n}{(1+r)^{n+1}}. \qquad (5)$$

Note that the capacity charge is applied to the new contracts only. The time to collect the capacity charge on the expired contracts was when the contract was written. In defining the capacity charge in this way, the reinsurer is making its desired rate of return on its allocated cost of capital[3].

We finish this section with the derivation of Equation 5.

- The reinsurer puts up an initial investment of $HM_0 \cdot \Delta C_0$ at time $t = 0$. It commits to holding $PRM_1 \cdot \Delta C_1$ at time $t = 1$, $HM_2 \cdot \Delta C_2$ at time $t = 2$, and so on.
- While the reinsurer is holding the capital, it is earning interest at rate i. At time $t = 1$, it expects to receive $HM_0 \cdot \Delta C_0 \cdot (1 + i) - HM_1 \cdot \Delta C_1$. At time $t = 2$, it expects to receive $HM_1 \cdot \Delta C_1 \cdot (1 + i) - HM_2 \cdot \Delta C_2$, and so on.
- The capacity charge is equal to the initial investment less the present value (at interest rate r) of the expected amount received. That is:

[3] One possible enhancement of Equation 5 would be to vary the rate of return by the length of time it is invested.

$$\Delta P = HM_0 \cdot \Delta C_0 - \sum_{n=0}^{\infty} \frac{HM_n \cdot \Delta C_n \cdot (1+i) - HM_{n+1} \cdot \Delta C_{n+1}}{(1+r)^{n+1}}. \qquad (6)$$

- Equation 5 is derived from Equation 6 by rearranging and grouping the terms in increasing order of n.

5. Examples

We now illustrate the use of our model and methodology on a number of sample reinsurance contracts. We constructed a reference portfolio of reinsurance contracts from real insurance companies, based on publicly available data. The lines of business in the reference portfolio included general liability, commercial auto, workers' compensation, professional liability, commercial multi-peril, fire, allied lines, earthquake and some personal lines. We treated the hurricane and earthquake exposures as separate contracts that took a 25% share of the underlying catastrophe reinsurance contract.

What follows is a description of the steps we took to construct this reference portfolio.

1. We first estimated the expected direct losses, by annual statement line of business, for the insurers included in the reference portfolio. For the most recent accident year, we estimated the expected losses by multiplying the reported premiums by our estimated loss ratio for the industry. For prior accident years, we used the insurers' reported loss reserves.

2. Using details of each insurer's reinsurance program reported to the A.M. Best Company, and the loss distributions underlying the ISO Underwriting Risk Model and the AIR catastrophe model for hurricanes and earthquakes, we partitioned the insurer's expected direct losses into two segments by

annual statement line of business – the expected net losses and the expected reinsured losses.

3. In order to project the expected release of the marginal capital over the next several years using Equation 5, we need to know the marginal capital attributed to current contracts in future accident years. This will depend on the reinsurer's business plan. We assumed that the reinsurer would continue its current business plan but, going forward, we estimated the expected unpaid losses using an ISO industry loss reserve study.

For the reference portfolio, the total expected loss for all lines in the current year is $739,998,127. The expected payout for losses from prior accident years is $1,813,101,644. In constructing this reference portfolio, we did not have all the detailed contract level information that is potentially available to a reinsurer. The reference portfolio had a few hundred contracts covering a variety of limits. The insurers in the reference portfolio tended to be larger than average and thus we expect the size and the retentions of the contracts to be a bit higher than normal.

We now describe how we calculated the necessary capital for the reference portfolio, with and without the proposed contracts.

1. In evaluating the non-catastrophe exposure we used the expected loss estimates and the limits for each contract. Using claim severity distributions in the ISO Underwriting Risk Model, we obtained the expected claim count by dividing the expected loss by the expected claim severity. The claim count distribution requires a second parameter that ISO obtained from analyses similar to that described in the appendix.

2. Using exposures that primary insurers reported to ISO, we ran the AIR catastrophe model to produce 10,000 simulated years of hurricane and

earthquake losses for each primary insurer in the reference portfolio. The catastrophe losses were adjusted to reflect the reinsurance provisions and the 25% share of the catastrophe contracts taken by the reference portfolio. The losses for all the catastrophe contracts in the reference portfolio were summed by year to produce a combined catastrophe size of loss distribution.

3. The distributional information above was used to derive the reference and marginal aggregate loss distributions by a procedure mathematically equivalent to Simulation Algorithm 3 above. Table 1 describes the aggregate loss distribution for the reference portfolio.

4. Following Equation 1, we set the needed capital for the reference portfolio equal to $TVaR_{99\%}(X) - E[X] = \$670{,}997{,}012$.

The next step was to calculate the heterogeneity multiplier, HM_n, for each year. This is done by finding the marginal capital for each reinsurance contract in the reference portfolio and applying Equation 2. While the heterogeneity multipliers varied slightly by year, they were all close to 1.64, well above the theoretical minimum of 1.00. Since we were assuming a stable business plan, we selected $HM_n = 1.64$ for all n.

Now we are ready to calculate the capacity charges for prospective reinsurance contracts using Equation 5.

The first set of examples consists of some standard property and casualty reinsurance contracts. We first calculate the marginal capital for the prospective contract for the current year and up to the following six years, which we assumed will have uncertainty in the ultimate paid losses. In this example, we are ignoring all uncertainty in ultimate losses after two years for Fire, after five years for

Commercial Auto, and after seven years for General Liability. In Table 2 we provide an illustrative aggregate loss distribution when a General Liability reinsurance contract are added to the reference portfolio. Table 3 gives the result of marginal capital calculations for the remaining contracts in this set of examples.

We used $HM_n = 1.64$, $r = 18\%$ and $i = 6\%$. The capacity charges calculated using Equation 5 for this set of examples are in Table 4.

Table 1

Aggregate Loss Distribution
Produced by
ISO Underwriting Risk Model
Reference Portfolio – Year 1

Aggregate Mean	2,553,099,771
Aggregate Std. Dev	226,983,918

Aggregate Loss	Cumulative Probability	Tail Value at Risk	Implied Capital
2,544,328,941	0.50000	2,733,373,669	180,273,898
2,572,708,866	0.55000	2,752,806,019	199,706,248
2,601,822,082	0.60000	2,773,508,440	220,408,669
2,632,192,534	0.65000	2,795,884,342	242,784,571
2,664,579,556	0.70000	2,820,498,110	267,398,339
2,699,943,781	0.75000	2,848,200,965	295,101,194
2,739,710,696	0.80000	2,880,411,932	327,312,161
2,787,036,572	0.85000	2,919,645,004	366,545,233
2,847,436,074	0.90000	2,971,590,416	418,490,645
2,887,426,613	0.92500	3,006,495,925	453,396,154
2,940,100,948	0.95000	3,053,590,102	500,490,331
2,953,219,034	0.95500	3,065,480,084	512,380,313
2,967,653,154	0.96000	3,078,622,916	525,523,146
2,983,799,371	0.96500	3,093,339,811	540,240,040
3,002,061,116	0.97000	3,110,102,957	557,003,187
3,023,031,771	0.97500	3,129,649,532	576,549,761
3,048,079,271	0.98000	3,153,220,036	600,120,266
3,080,209,479	0.98500	3,183,009,313	629,909,542
3,123,377,033	0.99000	3,224,096,783	670,997,012
3,195,198,671	0.99500	3,292,456,190	739,356,419
3,350,378,069	0.99900	3,446,040,482	892,940,711
3,416,123,232	0.99950	3,512,255,729	959,155,958
3,567,277,277	0.99990	3,670,239,190	1,117,139,419

Table 2

Aggregate Loss Distribution
Produced by
ISO Underwriting Risk Model
Reference Portfolio + General Liability Treaty B – Year 1

Aggregate Mean 2,554,099,777
Aggregate Std. Dev 227,010,259

Aggregate Loss	Cumulative Probability	Tail Value at Risk	Implied Capital
2,545,328,752	0.50000	2,734,394,860	180,295,083
2,573,712,090	0.55000	2,753,829,391	199,729,614
2,602,827,651	0.60000	2,774,534,227	220,434,450
2,633,202,873	0.65000	2,796,912,584	242,812,807
2,665,594,329	0.70000	2,821,528,848	267,429,071
2,700,955,923	0.75000	2,849,235,353	295,135,576
2,740,735,861	0.80000	2,881,449,351	327,349,574
2,788,066,006	0.85000	2,920,686,432	366,586,655
2,848,476,835	0.90000	2,972,637,184	418,537,407
2,888,468,111	0.92500	3,007,545,936	453,446,159
2,941,144,453	0.95000	3,054,644,769	500,544,992
2,954,263,865	0.95500	3,066,535,955	512,436,178
2,968,702,964	0.96000	3,079,679,982	525,580,205
2,984,851,364	0.96500	3,094,397,423	540,297,646
3,003,106,105	0.97000	3,111,162,780	557,063,003
3,024,076,687	0.97500	3,130,710,903	576,611,126
3,049,146,274	0.98000	3,154,284,003	600,184,226
3,081,283,973	0.98500	3,184,073,593	629,973,816
3,124,409,548	0.99000	3,225,162,418	671,062,641
3,196,271,436	0.99500	3,293,544,811	739,445,034
3,351,471,608	0.99900	3,447,105,396	893,005,619
3,417,154,542	0.99950	3,513,336,592	959,236,815
3,568,402,012	0.99990	3,671,307,958	1,117,208,181

Table 3

Reinsurance Contract	Marginal Capital Needed at Beginning of Year						
	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Fire A	52,488	11,869					
Fire B	54,694	12,428					
Fire C	66,358	15,383					
Comm Auto Liab A	34,962	28,845	18,913	16,435	7,944		
Comm Auto Liab B	37,350	30,810	18,045	15,255	7,533		
Comm Auto Liab C	52,799	44,260	27,308	19,896	9,560		
Comm Auto Liab D	40,810	33,850	21,319	16,976	8,064		
General Liability A	63,628	53,837	44,341	38,707	22,441	15,493	12,034
General Liability B	65,629	55,336	45,939	39,968	24,076	17,144	13,550
General Liability C	77,826	65,733	55,518	49,518	33,768	25,682	20,273
General Liability D	67,488	56,882	47,205	41,727	25,945	18,759	14,742

Table 4

Contract	First Contract		Second Contract		Expected Loss	Capacity Charge	Cap Chg as % Exp Loss
	Retention	Limit	Retention	Limit			
Fire A	500,000	500,000	–	–	1,000,000	10,432	1.04%
Fire B	1,000,000	1,000,000	–	–	1,000,000	10,878	1.09%
Fire C	1,000,000	5,000,000	–	–	1,000,000	13,241	1.32%
Comm Auto Liab A	500,000	500,000	–	–	1,000,000	14,525	1.45%
Comm Auto Liab B	1,000,000	1,000,000	–	–	1,000,000	14,942	1.49%
Comm Auto Liab C	1,000,000	5,000,000	–	–	1,000,000	21,174	2.12%
Comm Auto Liab D	500,000	500,000	2,000,000	2,000,000	1,000,000	16,561	1.66%
General Liability A	500,000	500,000	–	–	1,000,000	31,265	3.13%
General Liability B	1,000,000	1,000,000	–	–	1,000,000	32,484	3.25%
General Liability C	1,000,000	5,000,000	–	–	1,000,000	39,976	4.00%
General Liability D	500,000	500,000	2,000,000	2,000,000	1,000,000	33,695	3.37%

When you examine Tables 3 and 4, we hope you would agree that the capacity charges follow a logical progression in terms of relative risk and the length of time that capital must be held to support that risk.

Next we consider a set of examples consisting of catastrophe treaties. As we did in constructing the reference portfolio, using exposures that primary insurers reported to ISO, we ran the AIR catastrophe model to produce 10,000 simulated years of hurricane and earthquake losses for a number of primary insurers. The catastrophe losses were adjusted to reflect the reinsurance provisions, and we continued with the 25% quota share provision that was taken by the reference portfolio. For each contract, the losses were added to the losses of the catastrophe contracts in the reference portfolio by year to produce a combined catastrophe size of loss distribution. We assumed that there was no uncertainty in the catastrophe losses after one year.

Table 5 gives an illustrative aggregate loss distribution when a catastrophe contract is added to the reference portfolio. Table 6 gives the results of the marginal capital calculations and resulting capacity charges for each of the catastrophe contracts.

Table 5

Aggregate Loss Distribution
Produced by
ISO Underwriting Risk Model
Reference Portfolio + Earthquake C

Aggregate Mean	2,559,254,438
Aggregate Std. Dev	230,864,879

Aggregate Loss	Cumulative Probability	Tail Value at Risk	Implied Capital
2,549,206,753	0.50000	2,742,159,379	182,904,941
2,577,918,536	0.55000	2,762,007,818	202,753,380
2,607,399,694	0.60000	2,783,186,946	223,932,508
2,638,206,921	0.65000	2,806,117,885	246,863,447
2,671,104,813	0.70000	2,831,392,904	272,138,466
2,707,046,998	0.75000	2,859,915,077	300,660,639
2,747,671,983	0.80000	2,893,170,794	333,916,356
2,796,075,394	0.85000	2,933,840,121	374,585,683
2,858,223,270	0.90000	2,987,982,052	428,727,614
2,899,467,753	0.92500	3,024,575,910	465,321,472
2,954,132,749	0.95000	3,074,238,226	514,983,788
2,967,817,054	0.95500	3,086,832,661	527,578,223
2,982,939,047	0.96000	3,100,778,128	541,523,689
2,999,816,143	0.96500	3,116,425,597	557,171,159
3,018,878,770	0.97000	3,134,301,652	575,047,214
3,040,922,350	0.97500	3,155,211,119	595,956,681
3,067,778,750	0.98000	3,180,494,563	621,240,125
3,101,546,799	0.98500	3,212,599,452	653,345,014
3,146,195,429	0.99000	3,257,314,564	698,060,126
3,224,424,972	0.99500	3,331,993,080	772,738,642
3,396,310,134	0.99900	3,503,462,403	944,207,965
3,468,506,073	0.99950	3,578,282,216	1,019,027,778
3,642,049,455	0.99990	3,762,498,626	1,203,244,188

Table 6

Reinsurance Contract	Marginal Capital	Expected Loss	Capacity Charge	Cap Chg as % Exp Loss
Earthquake A	14,736	5,287	2,458	46.48%
Earthquake B	7,538,096	4,939,820	1,257,201	25.45%
Earthquake C	27,063,114	6,154,667	4,513,577	73.34%
Earthquake D	1,483,536	1,273,219	247,424	19.43%
Earthquake E	1,862,063	303,947	310,554	102.17%
Earthquake F	3,174,465	593,735	529,436	89.17%
Earthquake G	5,513,907	2,760,151	919,608	33.32%
Earthquake H	2,102,509	371,200	350,656	94.47%
Hurricane A	2,092,047	123,008	348,911	283.65%
Hurricane B	95,297	75,723	15,894	20.99%
Hurricane C	3,532,354	640,824	589,125	91.93%
Hurricane D	1,838,135	462,064	306,564	66.35%
Hurricane E	1,716,522	266,411	286,281	107.46%
Hurricane F	4,063,674	226,776	677,738	298.86%
Hurricane G	2,871,555	577,426	478,917	82.94%
Hurricane H	33,428,704	7,840,572	5,575,228	71.11%
Hurricane I	22,259,834	2,197,780	3,712,488	168.92%
Hurricane J	8,167,187	2,695,188	1,362,121	50.54%

A noteworthy feature of this last set of examples is the wide range of capacity charges. Earthquake Contracts E and G provide one of the nicer illustrations of what drives these differences. Table 7 gives some key statistics.

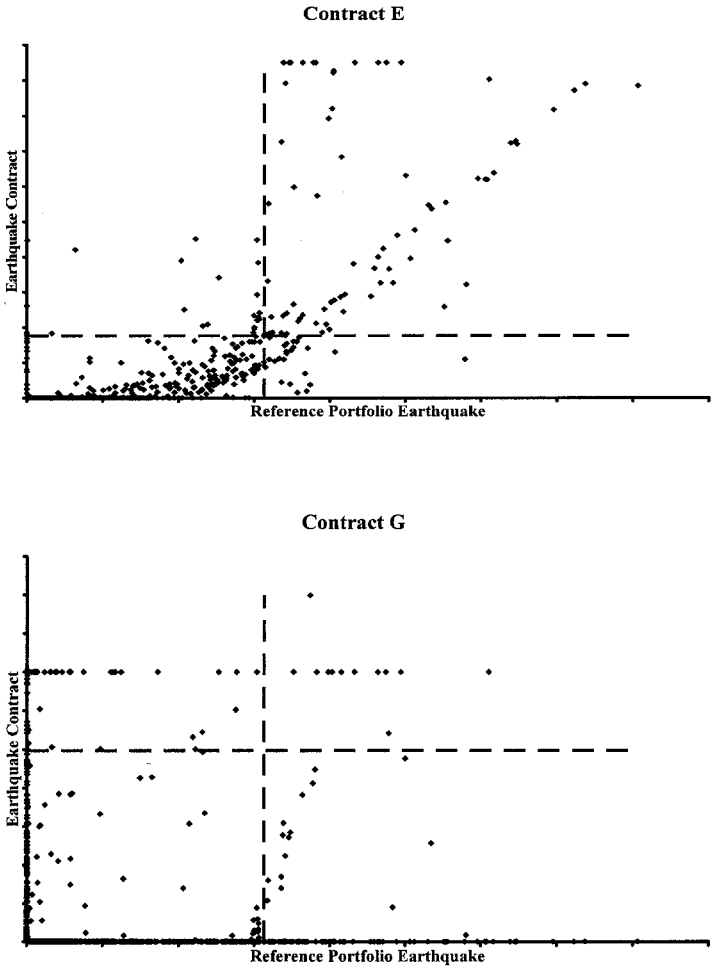
Table 7

Earthquake Contract	**Coefficient of Variation**	**Correlation with Reference Portfolio**
E	36.3	0.76
G	27.5	0.16

Earthquake Contract E is more volatile than Earthquake G, but the main difference is the correlation between the contracts and the reference portfolio. Figure 7 provides scatter plots of the contracts and the reference portfolio.

Figure 7

Note: The vertical and horizontal dotted lines represent the respective 99th percentile of the reference portfolio and the indicated reinsurance contract for earthquake reinsurance.



7. Summary and Conclusions

This paper started with three objectives:

1. Demonstrate a practical method to determine the distribution of a reinsurer's aggregate loss payments.

2. Using the results of Objective #1, demonstrate how to determine the amount of capital needed for a reinsurance company based on its risk of loss.

3. Using the results of Objective #2 demonstrate how to determine the capacity charge for a new reinsurance contract.

We demonstrated our methodology for accomplishing these objectives on an illustrative reinsurer with hundreds of reinsurance contracts.

We used the ISO Underwriting Risk Model to determine the aggregate loss distribution. As input, the model took the limits and quota share percentages for each reinsurance contract for the "standard lines" of insurance. 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.

Next we determined the capital needed for the reinsurer by calculating the Tail Value-at-Risk from the aggregate loss distribution.

Finally we calculated capacity charges for a variety of reinsurance contracts. The rationale underlying these calculations was that the total capacity charge over all

reinsurance contracts should provide the reinsurer with a competitive expected return on capital.

The underwriting strategy used to get this expected return assumed that the reinsurer will write those contracts that provide the greatest return on marginal capital. Now it can take several years for some reinsurance contracts to be settled. The reinsurer must hold capital as long as there is uncertainty in the final settlement of its claims, and the capacity charge reflects how long capital must be held because it reinsures a given contract.

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

8. Additional Comment

There is recent actuarial literature on "correlation in the tails" such as that of Venter [2002]. The analysis documented in the appendix of this paper estimates an overall level of correlation not attributed to particular region of the loss distribution. We doubt that we have sufficient data to make such an attribution.

Furthermore, to the extent that correlation in the tails is driven by large natural catastrophes, we argue that, when we couple a collective risk model parameterized by the parameters estimated in the appendix with simulation runs from a catastrophe model, as documented above, we do indeed capture at least some "correlation in the tails."

Should a reinsurer want to use a copula, or some other dependency model, our methodology for determining the needed capital and capacity charges can accommodate it. At the very least, one can generate a large number of stochastic scenarios and incorporate that into the collective risk model in exactly the same

way that we did for the catastrophe model.

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.

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. Gary Venter, "Tails of Copulas," *Proceeding of the Casualty Actuarial Society LXXXIX*, 2002. http://www.cas.org/library/02pcas/venter.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 reinsurance 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 $(.03)^2 \approx 0.001$. 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]:

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

- $\delta_{ii'}$ 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:

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

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 $i \neq i'$ or $j \neq j'$. Products in which the first and second LOBs are the same, $i = i'$, but companies are different, $j \neq j'$, yield estimators for the covariance generators g_i. Products in which the first and second LOBs are different, $i \neq i'$, 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 95^{th} or 99^{th}) 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.

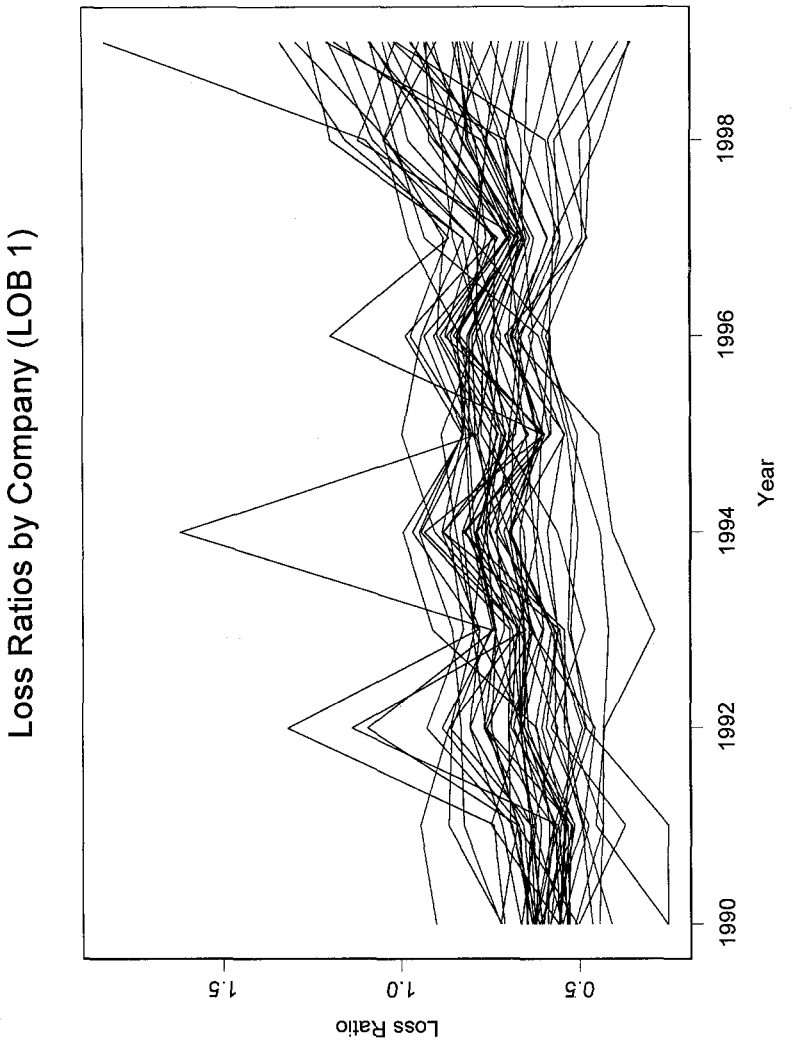
Considering next pairwise products with first and second LOB equal to LOB 1 and with first and second companies equal ("within company"; in other words, the squared deviation products) yields an estimate for LOB 1 of contagion plus covariance generator of 0.0226 with a standard error of 0.0092. This 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.

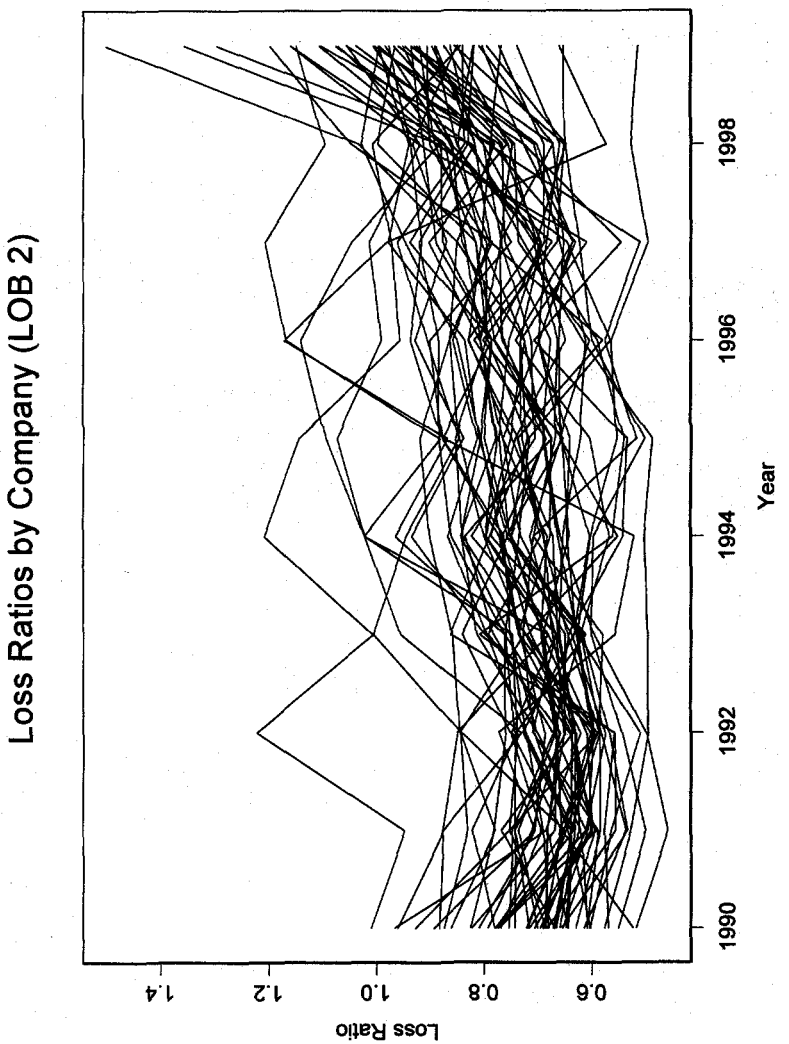
Exhibit A.2 shows the same statistics for LOB 2, a g estimate of 0.0007 with standard error of 0.0004 (hence just about significant at two standard errors, indicating a frequency parameter uncertainty of plus or minus 2.6%), no significant size dependence of this g estimate, and a significant estimate 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.

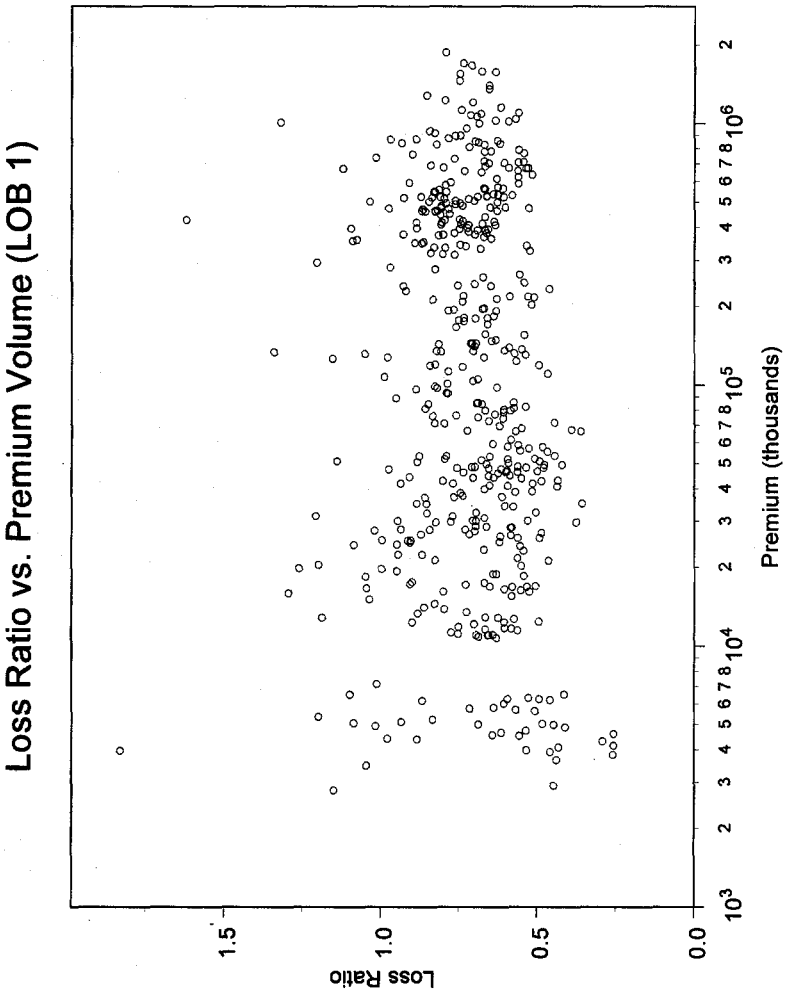
Graph A.1.1


Loss Ratios by Company (LOB 1)
Loss Ratio
0.5
1.0
1.5
1990
1992
1994
1996
1998
Year

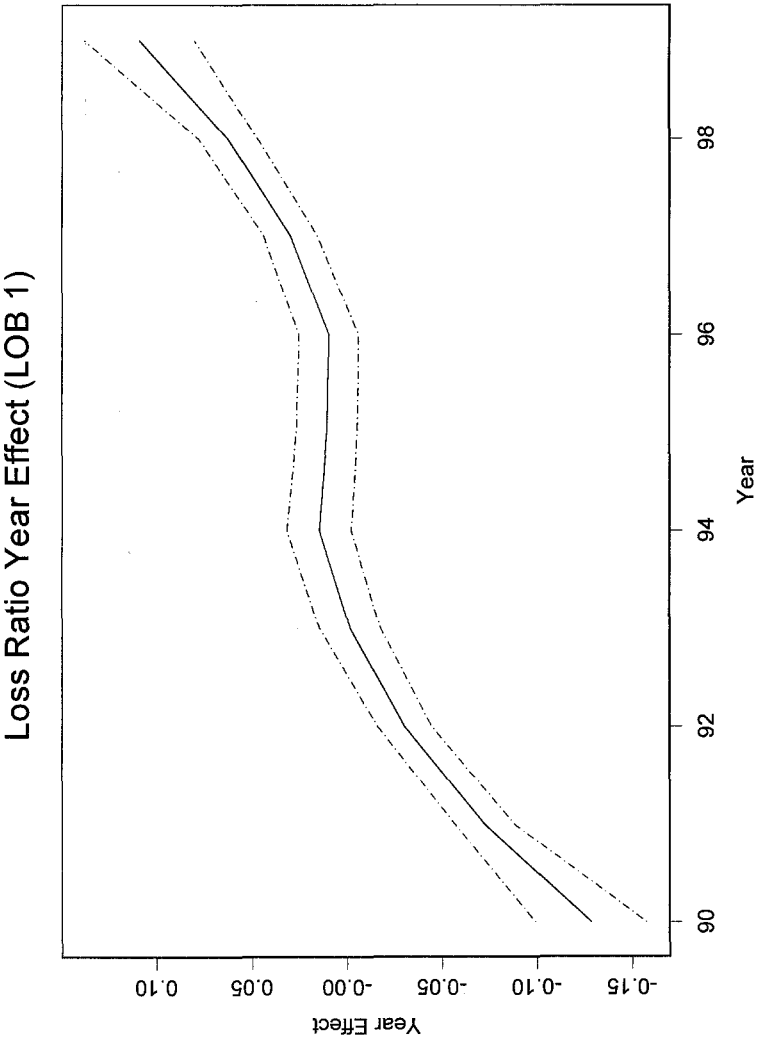
Graph A.1.2


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

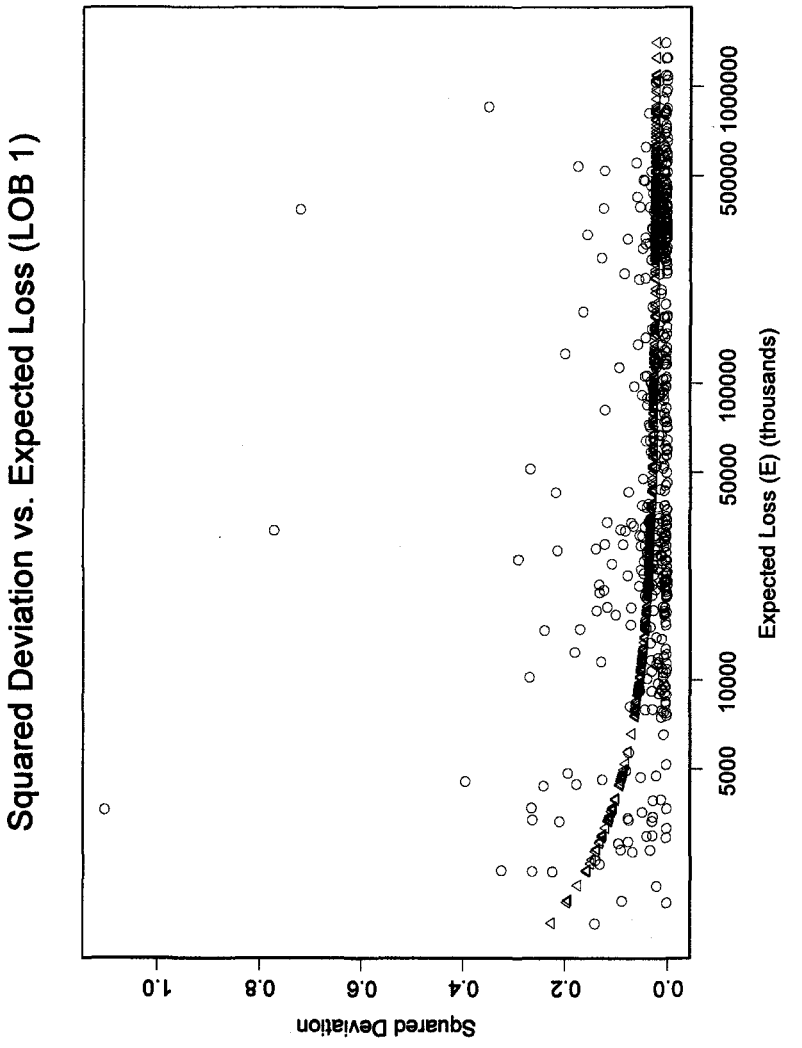
Graph A.2


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

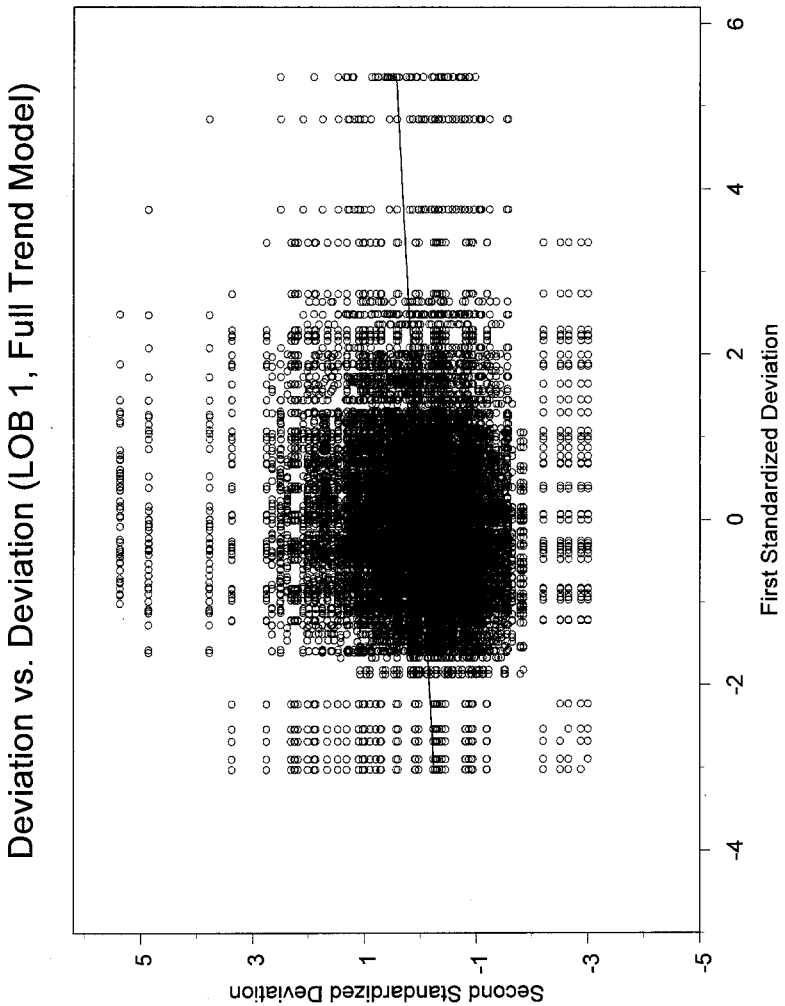
Graph A.3


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

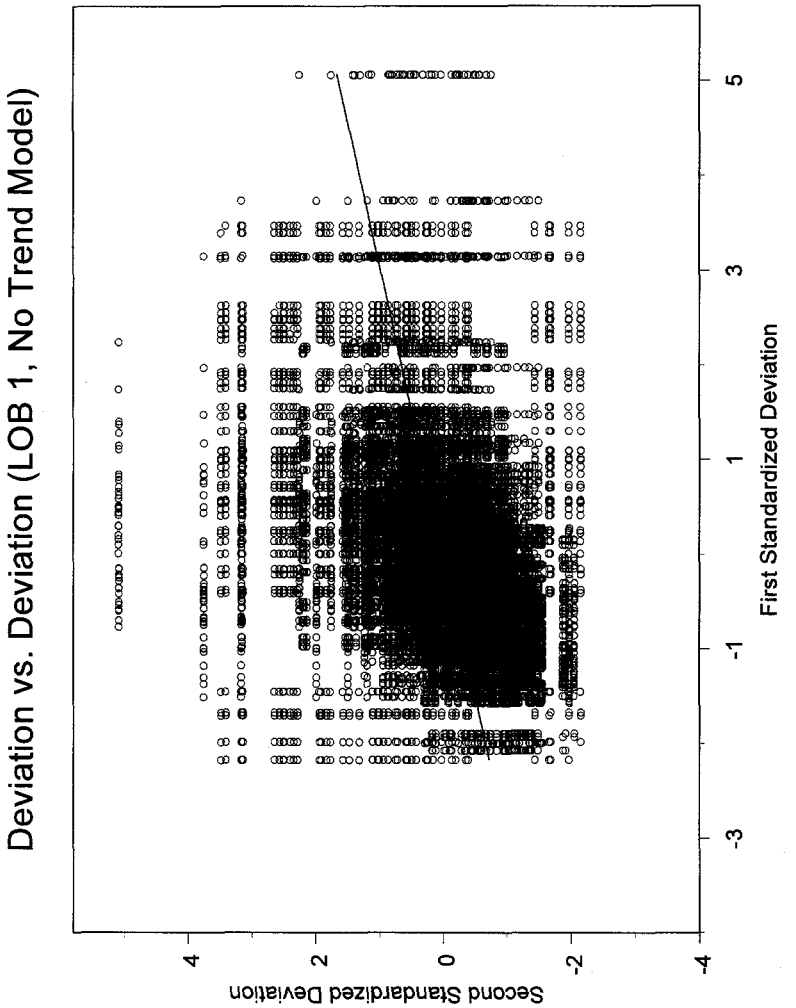
Graph A.4


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)

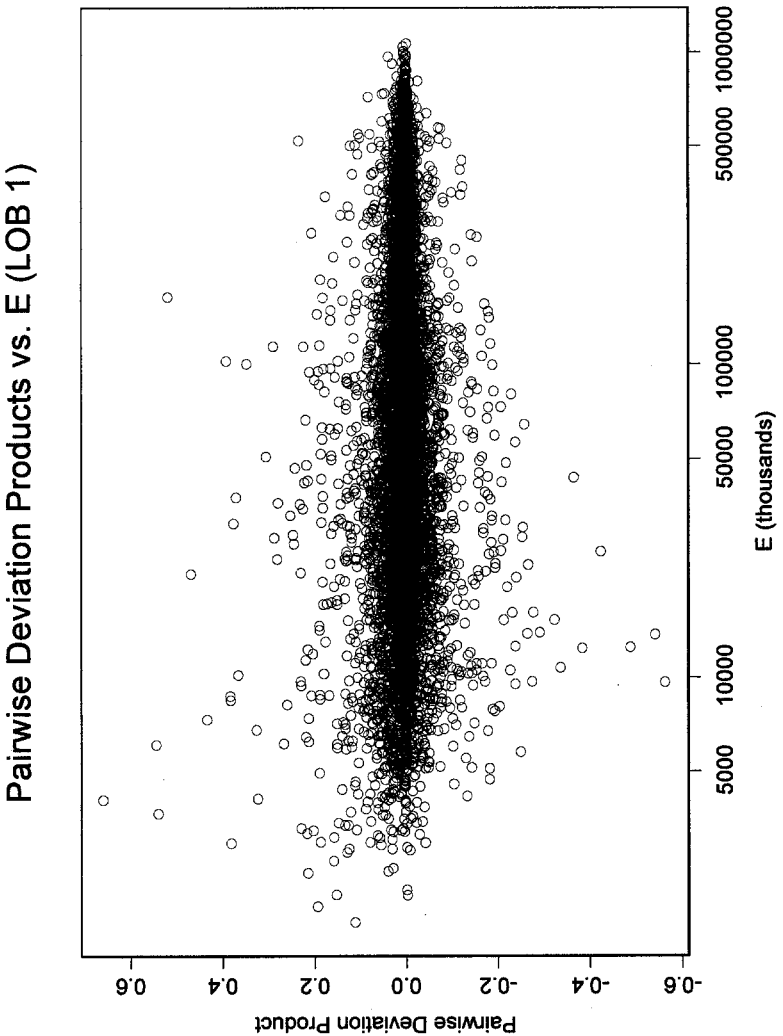
Graph A.5.1


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

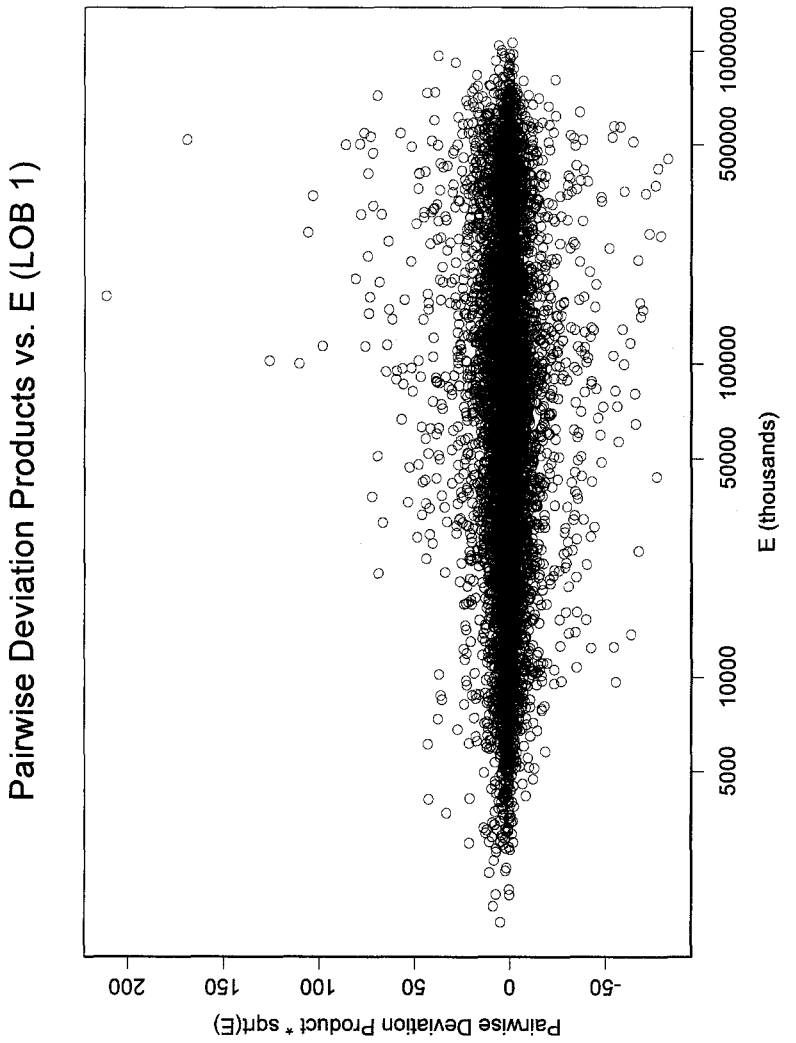
Graph A.5.2


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

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

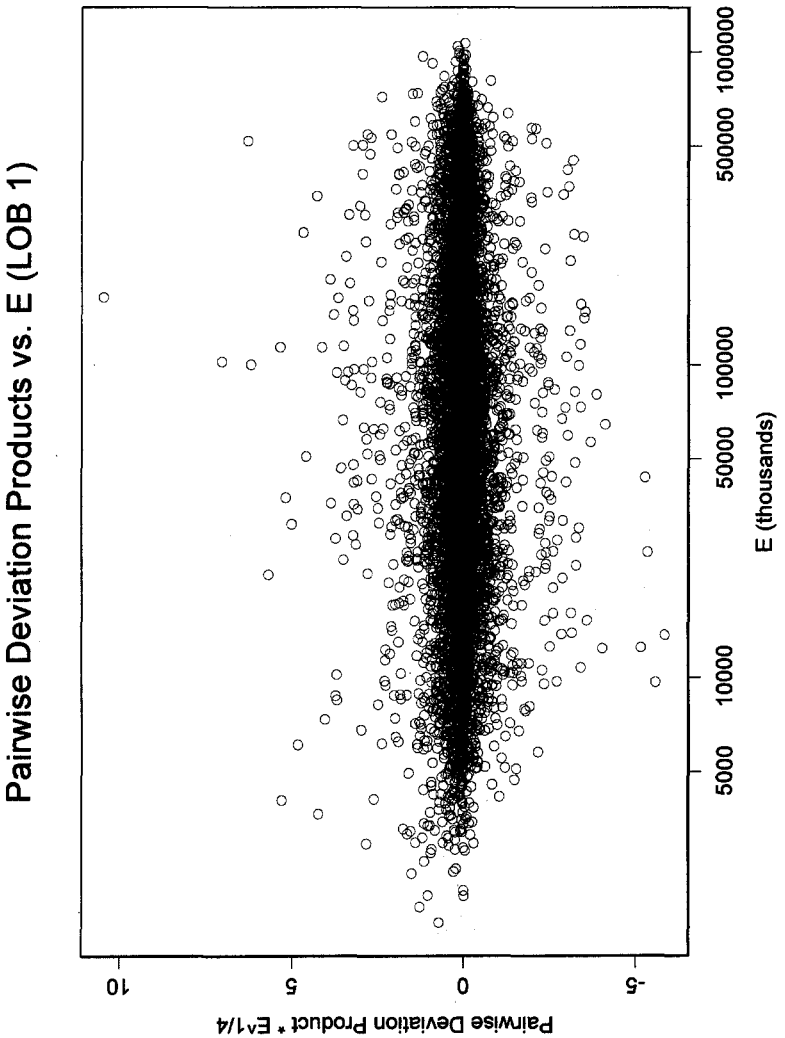

Pairwise Deviation Products vs. E (LOB 1)
Pairwise Deviation Product * E^1/4
10
5
0
-5
5000
10000
50000
100000
500000
1000000
E (thousands)

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

Correlation

Thomas Struppeck, FCAS, MAAA

Correlation

Thomas Struppeck, FCAS, MAAA

Abstract

Actuaries frequently are called upon to estimate sums of random variables. Such sums arise in a variety of contexts, as aggregate loss distributions, as losses including loss adjustment expense, as losses to a particular layer in stop loss reinsurance. If the quantities being summed were independent, things would be easy, however this is seldom the case. Generally, there will be some amount of "correlation" between the summands.

This paper examines the Pearson product moment correlation coefficient's strengths and weaknesses and discusses two non-parametric alternatives: the Spearman rank correlation coefficient and Kendall's tau statistic.

Correlation

Cor-re-la-tion n.

1. A causal, complementary, parallel, or reciprocal relationship, especially a structural, functional, or qualitative correspondence between two comparable entities.
2. Statistics. The simultaneous change in value of two numerically valued random variables.[1]

What is correlation? As can be seen from the above dictionary definition, correlation is a difficult concept to precisely define. Complicating the matter is the fact that the word "correlation" has a common usage (definition 1) and a technical usage. Upon hearing the word "correlation", actuaries may think of the (Pearson product moment) correlation coefficient, this preconceived notion can lead to further confusion.

In this paper, we will consider several notions of correlation and examine their relative strengths and weaknesses. After this brief introductory section, we will develop the Pearson product moment correlation coefficient and see why it specifically is central to the problem of estimating aggregate loss distributions. In the third section, we examine two other possible measures of correlation, Spearman rank correlation and Kendall's tau. These measures are in some ways more natural than product moment correlation, and they have appeared at various places in the actuarial literature.

In the fourth section, we examine some ways of estimating correlation coefficients. As we shall see, this can be a very tricky business. Sometimes pairs of quantities that at first glance would appear to be highly correlated, turn out not to be. We will see an example in which two independent random variables appear to be correlated!

Oftentimes the goal of studying some correlated random variables is to obtain an aggregate distribution. A common method for estimating the aggregate distribution is to use Monte Carlo methods. Our fifth section is devoted to examining different ways of simulating correlated random variables along with some cautions for the various methods.

---

[1] *The American Heritage® Dictionary of the English Language, Fourth Edition, 2000, Houghton Mifflin Company.*

Finally, before the Conclusions section, we consider a hypothetical aggregate stop loss and examine how our choice of correlation model can vastly change the estimated value of the different layers.

Independence and Correlation

In the normal course of business insurers commonly extend vastly more in total limit outstanding than they have in total assets. Consider a homeowners writer that insures 10,000 houses in each of fifty states for $100,000 each. The total sum insured is $50 billion. Assuming a premium of $1,000 per house, total annual premiums would be $500 million, so policyholder surplus might be $200 million. In this example, the total sum insured is 250 times the available surplus. How can such leverage work? The answer, of course, is that the insurer does not expect claims from all of its insureds at once. There is an assumption that losses occur somewhat independently from one another. This spreading of the risk is the essence of insurance.

What is Correlation?

First, let us review some probability. Suppose that X and Y are random variables with finite means, then:

$$E(X+Y) = E(X) + E(Y).$$

That is, the expected value of the sum is the sum of the expected values. Since expected value is the amount that the insurer will have to pay on average, it is naturally of considerable interest to insurers.

Insurers are also concerned about the variation around this expected value. This variability is commonly measured by using the standard deviation or its square, the variance. The variance is the expected squared deviation from the mean, namely:

$$VAR(X) = E((X - E(X))^2)$$

Or equivalently:

$$VAR(X) = E(X^2) - E(X)^2$$

Unlike the rule for expected value, the variance of a sum is not sum of the variances, except under specific conditions. Instead there are other terms:

$$\begin{aligned} VAR(X+Y) &= E((X+Y)^2) - E(X+Y)^2 \\ &= E(X^2 + 2XY + Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2E(XY) - 2E(X)E(Y) \end{aligned}$$

Denoting $E(XY) - E(X)E(Y)$ by $COV(X,Y)$, we obtain:

$$VAR(X+Y) = VAR(X) + VAR(Y) + 2\,COV(X,Y)$$

As we see above, computing the variance of the sum of X and Y leads us to consider the expected value of the product of X and Y and the product of their expected values. This product term in turn motivates the definition of covariance. It is worth noting that the covariance of X with itself is the variance of X.

Covariance is one step away from our goal, the correlation coefficient. If X is expressed in one unit of measure (say, meters) and Y is expressed in another unit (say, seconds), then the unit for the covariance will be the product (in this case meter-seconds). It is possible to normalize the covariance (assuming finite, non-zero second moments for X and Y), by dividing by the product of the standard deviations of the two variables. In his delightful book [F], Feller[2] suggests that a physicist might call this quantity "dimensionless covariance"; this ratio is the Pearson product-moment correlation coefficient. Since we will be looking at several possible measures of correlation, we will henceforth refer to this measure as "dimensionless covariance".

This quantity measures the extent to which X and Y are linearly related. It ranges between minus one and one. Being a measure of the linear relationship between two quantities might lead one to suspect that there is some connection between the correlation coefficient and linear regression. Indeed there is such a relationship, and it highlights a subtle point. Select a sample from the jointly distributed population (X, Y). From this sample, first compute the sample correlation coefficient. Then do two regressions: regress Y on X, and X on Y. In each of these regressions, X and Y have asymmetric roles: X is the independent variable and Y is the dependent

[2] Every actuary should read, at a minimum, section I.5 of Volume Two, "The Persistence of Bad Luck".

variable in the first and vice versa for the second. From each of the two regression lines, a slope is obtained. The sample correlation coefficient is the geometric mean of these two slopes (see for example, [CS].) The sample correlation coefficient treats X and Y equally; neither is thought of as being a function of the other, they are simply related somehow.

There are several ways that the correlation coefficient could be zero. The most important case is when the variables are independent. Independence, however, is a much stronger condition than correlation zero. Recall that two or more random variables are called independent if their joint density function factors into the product of their respective density functions. If two random variables are independent, knowledge about the value of one of them tells you nothing new about the value of the other one; i.e. the conditional distribution is with probability one the same as the unconditional distribution. On the other hand, if two random variables have a non-zero correlation, then knowing the value of one of them might give you information about the other one.[3]

Here is an example. Consider two lines of insurance, A and B, which have losses jointly distributed as follows:

	1	2
1	1/6	1/6
2	1/4	1/12
3	1/12	1/4

So the losses for line A can be either 1 or 2 and the losses for line B can be 1, 2, or 3 with the probabilities for the 6 possible pairs of events given by the above table. This table is the joint distribution function and it contains all of the information about A and B. For example, by summing the columns we get the marginal distribution for A, namely 50% of the time it is 1 and 50% of the time it is 2. Similarly by summing the rows, we get the marginal distribution for B, which is also uniform, this time on the three values 1, 2, and 3.

If A and B were independent, the probability of any given pairing would be 1/6. Evidently A and B are not independent. If we compute the Pearson correlation coefficient, we find that it is about 20.4%. It is

[3] More exactly, the probability that the conditional distribution is different than the unconditional distribution is positive.

interesting to note that the conditional distribution of A given that B=1 is the same as the unconditional distribution of A; however, if we are given that B=2 instead, then we have a different conditional distribution for A. (A in this case has a 75% chance of being 1 and a 25% chance of being 2, as the reader may check.)

Correlation and covariance are central concepts in modern portfolio theory. Consider a simple portfolio consisting of N securities, X_i, each of which has a value that is randomly distributed with mean M and variance V. Further, suppose that each pair of security prices has covariance equal to C, then:

$$E(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} E(X_i) = N \cdot M$$

and:

$$VAR(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} VAR(X_i) + 2\sum_{j<i} Cov(X_i, X_j)$$

$$= N \cdot V + 2\binom{N}{2} C = N \cdot V + 2\frac{N(N-1)}{2} C$$

The first sum in the variance line has N terms, but the second sum has N(N-1)/2 terms. As N gets larger, the second term, which is quadratic in N, starts to dominate. What happens if we examine the average instead of the sum of the X_i?

$$VAR(\frac{1}{N}\sum_{i=1}^{N} X_i) = \frac{1}{N^2} VAR(\sum_{i=1}^{N} X_i)$$

$$= \frac{N \cdot V}{N^2} + 2\frac{N \cdot (N-1) \cdot C}{N^2 \cdot 2} = \frac{V}{N} + \frac{N^2 - N}{N^2} C$$

The first term goes to zero as N goes to infinity, but the second term is bounded away from zero by C, the pairwise covariance. This last term, called "non-diversifiable risk" (because it doesn't go away as N gets large), is a cornerstone of the capital asset pricing model (CAPM) and of other factor models such as arbitrage pricing theory (APT).

Knowing the correlation coefficient is not enough

There are many ways that two random variables can co-vary. Even if we know the marginal distribution of X and Y and we know the correlation of X and Y, there generally will be many possible joint distribution functions, and hence, different ways that they might co-vary. A good way to visualize the relationship between two variables is to examine a scatterplot of them.

Consider the two data sets in Figures I and II. Both have standard normal marginal distributions, and the each pair of variables has correlation coefficient $\rho = 80\%$. The pair of normal random variables shown in Figure I is multivariate normal. We will use the term "binormal" for a pair of multivariate normal random variables. Each of the random variables shown in Figure II is normal, but their joint distribution function is not the multivariate normal. The pair in Figure II has been engineered so that the random variables exhibit extreme behavior simultaneously much more often than the pair in Figure I, yet the pairs in Figures I and II both have $\rho = 80\%$. Notice how the second and fourth quadrants have fewer observations in Figure II than in Figure I. These two quadrants represent pairs of random variables that have opposite signs. Similarly, the first and third quadrants appear to have more observations in Figure II than in Figure I, especially away from the origin. These points represent observations with both random variables being exceptionally far away from average.[4] This comparison illustrates a major problem with using any single number to attempt to describe how two things co-vary: one number cannot do the job.

Remarkably, for continuous random variables, there is an object that does capture the exact way that two or more random variables co-vary. Furthermore, it captures only the interaction between the variables, being independent of their respective marginal distributions. This object is called a copula and is described briefly in a following section. The next section serves to give further motivation to the introduction of copulas.

[4] The probability of an observation being in quadrants one or three minus the probability of an observation being in quadrants two or four is one way to define Kendall's tau, which we will discuss later.

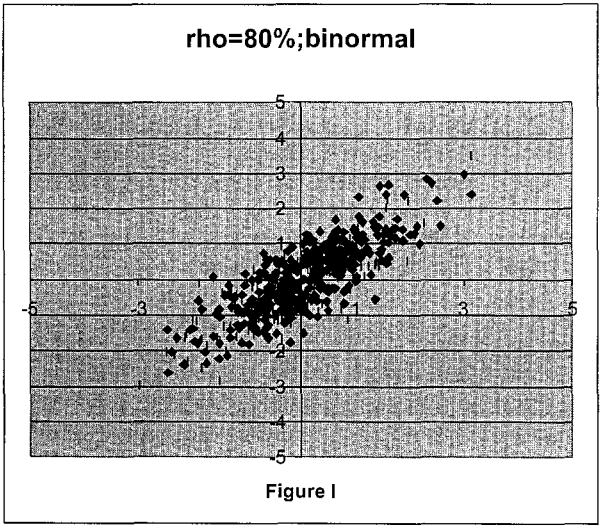


Figure I

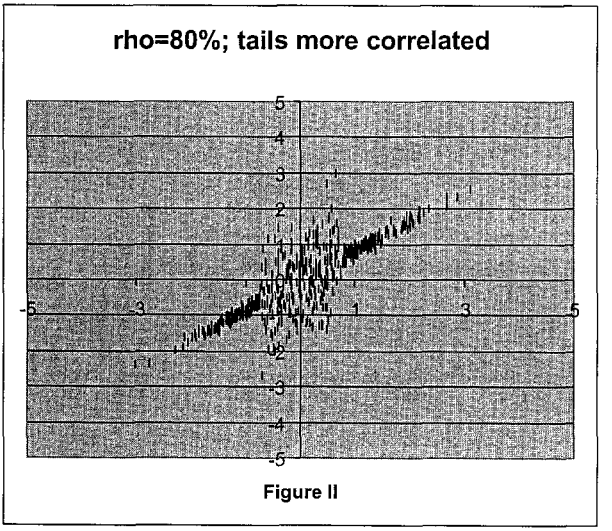


Figure II

Mango and Sandor's experiment

In a recent paper [MS], Mango and Sandor report on an experiment they performed to test a simulation procedure for generating correlated excess losses. They were attempting to estimate how much capital a reinsurer needs to support its book. They began by generating a binormally distributed pair of random variables with mean 0, variance 1, and specified correlation. As is well-known, a multinormal random variable is uniquely defined by its covariance matrix and its vector of means.

They take the pair (X, Y) of normal random variables and compute from it a pair (U, V) of lognormal random variables by setting U = exp(X) and V = exp(Y). They then censor these observations from below retaining only the pairs (U, V) where both exceed some threshold. At this point, they compute the correlation coefficient of the censored pair and observe that it is considerably smaller than the original correlation used when generating the pair (X, Y). Since the amount of correlation strongly influences how much capital is needed, they found this result to be alarming.

It will be instructive to examine these steps again, inserting one intermediate step.

Step 1: Generate a pair (X, Y) of normal random variables, binormally distributed with mean 0, variance 1, and specified dimensionless covariance.

Step 2: Denote by F() the cumulative distribution function for a normal random variable with mean 0 and variance 1. Determine the point in the unit square given by F(X), F(Y). In other words, determine the joint cumulative distribution function for this pair. (This step we have inserted into Mango and Sandor's process.)

Step 3: Compute U = exp(NORM(X)) and V = exp(NORM(Y)), where exp is the exponential function and NORM is the inverse of the cumulative distribution function that we introduced into the process in Step 2.

Step 4: Censor the results from below deleting all observations that have either U or V less than some threshold.

Step 1 is the generation of a pair random numbers and in this case the imposition of some sort of correlation structure on them.

Step 2 takes these random numbers and transforms them into a distribution function over the unit square. This particular choice of random number generation and transformation creates what is called a normal copula. The marginal distribution for each of the two coordinates is uniform on (0,1).

Step 3 takes these uniform (0,1) marginals and converts them to specified (in this case lognormal) marginals.

Step 4 takes the pair of lognormally distributed random variables and censors them from below.

Mango and Sandor noticed that the dimensionless covariance dropped. In which steps did they lose it? Some was lost in step 2 and some more in step 3 (they combined these steps). Also, some was lost in Step 4.

The loss in steps 2 and 3 comes about because dimensionless covariance is not preserved under non-linear transformations. A binormal pair of random variables is created in Step 1 and two nonlinear transformations are applied to it. There are measures which we will discuss later that are preserved under these transformations.

The loss in step 4 is more fundamental. The process of censoring is not a transformation --- some observations are discarded. Paraphrasing Embrechts, et al [EMS], the normal copula exhibits independence in the tails. What this means is that as we focus our attention on the observations in the extreme upper right portion of the first quadrant the pairs of observations that we find there are distributed as though they were (nearly) independent. This happens even though if we look at all of the observations they will exhibit some correlation.

Copulas

Recently the notion of a copula has begun to become more widely known by North American actuaries. Venter's paper, Tails of Copulas, is a particularly nice source [V]. (Other good sources are [W], [EMS], and Nelson's book [N].)

A copula is a dependence structure between two or more random variables. Let X and Y be (real-valued) random variables with joint distribution function F(x,y). That is:

$$F(x,y) = \Pr(X \leq x \text{ and } Y \leq y) \text{ for all real x and y}$$

If we just consider the variable X, we can ask about its distribution. It is given by the marginal distribution function, $F_X(x)$. Similarly, the marginal distribution of Y is given by $F_Y(y)$.

If we can find a function C: $[0,1]^2 \rightarrow [0,1]$ with $F(x,y) = C(F_X(x), F_Y(y))$, then we call that function a copula for X and Y.

Copulas have many nice features. First, they always exist. Secondly, in the case when X and Y are continuous, they are unique. This feature follows from what is known as Sklar's Theorem. It would take us too far afield to discuss this important result further. (The interested reader should see, for example, [ELM] or [N].)

One corollary of Sklar's Theorem is that, given a pair of continuous marginal distributions, every possible joint distribution with those marginals is given by a copula. The copula contains all of the information about the dependence structure of the joint distribution function; the marginal distributions contain the information about the individual components.

Considering the above, it seems that a good measure of the dependence of two (continuous) random variables would depend only on their copula since the copula captures all of the dependence information. Alas, dimensionless covariance is not a function of just the copula (in general it will depend of the marginals also), but there are measures that depend only on the copula. Since the copula captures all of the dependence information, this would suggest that these other measures are more natural. We discuss two of these measures later, but first let us see some other problems with dimensionless covariance.

Problems with dimensionless covariance

As we have seen, dimensionless covariance has several nice properties. It is easy to compute an estimate of it from a sample. It is related to linear regression, a familiar topic. In the case of a standard multivariate normal, knowing the pairwise correlations (equivalently, the covariance matrix) completely determines the joint distribution.

Ironically, these same properties highlight major weaknesses with it. The fact that it is easy to compute may lead practitioners to use it when it is not appropriate. The relationship to linear regression may make management feel that this is the right measure to consider, even if better

measures are offered. And finally, the fact that the covariance matrix contains all of the pertinent information in the multivariate case has led many to believe that it is true in general, when in fact, this relationship is seldom true.

The dimensionless covariance is invariant under positive affine transforms. I.e. if you multiply one of the variables by a positive constant and add another constant, the transformed variables have the same correlation coefficient. However, it is not preserved under more general transforms even if these transforms are monotone increasing. This noninvariance is a distinct weakness which seems to still not be fully understood by all practicing actuaries. Wang and Mango-Sandor touch on this issue.

Another problem with dimensionless covariance was alluded to earlier, namely that it is a single number. There are other problems too. While the dimensionless covariance is always between minus one and one, it is not always possible to achieve those bounds. Wang [W] gives explicit bounds for some lognormal random variables and shows that the range of achievable correlations can be made arbitrarily small. Wang shows that the largest possible dimensionless covariance between a lognormal random variable with parameters μ and 1 and another one with parameters $\mu \cdot \sigma$ and σ^2 is:

$$\frac{e^{\sigma} - 1}{\sqrt{e^{\sigma^2} - 1}\sqrt{e - 1}}$$

As the reader may check, this quantity tends to zero as sigma tends to infinity. So, in particular it is possible to have a pair of comonotonic lognormal random variables with dimensionless covariance arbitrarily close to zero.

Here is another easy example. Let X and Y be Bernoulli random variables with success probabilities of 10% and 20%, respectively. To maximize the correlation coefficient, we make X and Y equal to one simultaneously as often as possible. To do this we select a uniform (0,1) random variable, U and define:

X = 1, if U<0.1 X = 0, otherwise

Y = 1, if U<0.2 Y = 0, otherwise

We now compute the covariance of X and Y.

$$COV(X,Y) = E(XY) - E(X)E(Y) = 0.1 - 0.02 = 0.08$$

The standard deviations of X and Y are 0.3 and 0.4, respectively, so their correlation coefficient is 2/3. These two variables move together as much as possible, but their correlation coefficient is not 1. The reader can check that the smallest possible correlation between such a pair of Bernoulli random variables is $-1/6$.

In the case of multivariate normal random variables (or more generally elliptical distributions), the correlation matrix is the canonical measure, i.e. specifying the correlation matrix uniquely defines the distribution. Therefore, it may be surprising to some that the correlation lives near the center of the distribution. What I mean by that is that, in the tails, there is little correlation. As mentioned earlier, this is what Embrechts means when he says that the normal copula exhibits tail independence. Mango and Sandor observe this property in their study in which they consider lognormal random variables [MS].[5]

It is easy to produce a spreadsheet that generates, say, 500 pairs of standard binormal random variables with ρ=60% and then computes the sample correlation when these are truncated at 1 from below or censored at 1 from below. If you make such a spreadsheet, you will observe that in both cases, the resulting population correlation is smaller than 60% and the observed sample correlation is usually smaller. These examples illustrate two potential missteps for modelers:

Mistake 1: Assuming that by modeling the ground-up losses and their correlations accurately, you will automatically get the excess loss correlations correct.

Mistake 2: Assuming that because you know that the excess layers do not exhibit significant correlation, the ground-up losses will also not exhibit any.

Finally, correlation is very hard to measure. While the sample correlation is easily computed, the confidence intervals around it are quite wide, even under the assumption that the variables are binormally distributed. The author's spreadsheet exhibits considerable instability in the

[5] Since the tail independence is a property of the copula, one expects that they would have observed something similar for any marginal distributions.

sample correlations --- even in the sample correlations of the binormal random variables themselves. Brehm [B] alludes to some of the problems of measuring correlations from data.

Here is one suggestion: when analyzing data, looking at a scatterplot of the data is almost always a good idea.

Consider this pair of jointly distributed random variables:

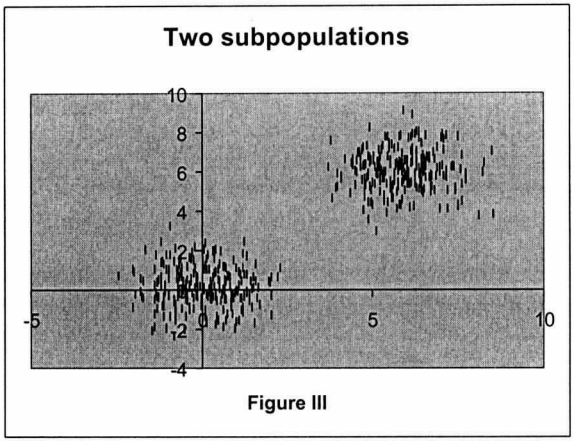


Figure III

A computation of the correlation coefficient will show a significant correlation, however it is misleading. As we see from examining the plot, we really have two classes. Once the two classes are identified and split, the apparent correlation goes away. We will see a second example of this type when we look at simulating correlated random variables. The variables here could be profitability for two lines of business, perhaps auto and general liability. One cluster might be urban risks and the other cluster rural risks. Given that a risk is rural, the auto and general liability are independent, similarly, given that the risk is urban, the auto and general liability are independent. However, naively computing the correlation of these two lines without partitioning the data, we will be misled into thinking that the lines are highly correlated.

Care should also be taken when adjusting data. Commonly, one adjusts for changing exposure bases over time by examining loss ratios. A danger with this is that loss ratios depend on premium and premium depends on the underwriting cycle. It is possible that apparent correlation between

losses from various lines is actually simply simultaneous fluctuations in premium adequacy caused by the underwriting cycle. If correlation between loss ratios is to be computed one should adjust for any premium redundancy or inadequacy first.

Another computational danger arises from binning of data. Insurance data frequently has a large degree of "lumpiness". Sometimes data is binned (perhaps all values between 100 and 300 are coded as 200). In addition, policy limits, deductibles, and case reserving methodologies can cause observations to be clustered around certain values. These data quirks can play havoc with canned correlation calculating software. Even if a more detailed analysis is attempted, the loss of detail in the data may lead to spurious results. The non-parametric methods described in the next section are more robust in this regard.

Since the computation of dimensionless covariance is essentially fitting a least squares line through the data, it is very sensitive to outliers. Observations that appear to be outliers should not just be ignored. Their existence should immediately call into question the assumption that the data is binormal (insurance data almost never is). There are non-parametric measures that are not as sensitive to the exact population distribution. We will examine two of these in the next section.

So, in summary, there are several problems with dimensionless covariance as a measure of correlation:

1) Confidence interval tests are usually predicated on the underlying data being binormal;
2) Insurance data frequently is binned or dirty in some other way, when reviewing this type of data, computing correlation is especially dubious;
3) Seldom can every possible value between -1 and 1 be achieved;
4) It is not preserved under increasing transformations; and finally,
5) It can be overly sensitive to outliers and to the distribution of the population which is almost never know.

Other Measures

As central to the study of aggregate losses as correlation and covariance are, as we shall see later, they really are not well behaved with respect to transformations of the data. There are other notions of correlation that are better behaved in that regard and are in some ways more intuitive than correlation.

The first of these measures is Spearman rank correlation. The idea behind rank correlation is very simple. Suppose that we have N pairs of observations of X and Y. (Remember that we are trying to understand how they are jointly distributed, so we must always think of the pair.) We look at the set of observed X values and the set of observed Y values (for the moment, assume no repeated values in either set). We take the N values of X and rank them from smallest (which gets assigned the number 1) up to largest (which gets assigned the number N). We similarly rank the Y values. We now look at the pairs of rankings, which we show sorted by X ranking:

Pairs (X, Y)	Pairs (Rankings)
(2.1, 7.4)	(1, 3)
(4.9, 9.8)	(2, 4)
(7.7, 7.3)	(3, 2)
(8.0, 5.2)	(4, 1)

Notice that all of the information about X and Y is gone, all that is left is information about how they are paired together. I.e. it is a function of their copula. With all of the information about X and Y abstracted away, we have reduced the problem to one involving only finite sets. Once we know N and the pairs of rankings, we have extracted all of the information.

Having sorted the pairs in order by their X rankings, we now count the number of "inversions" in the Y rankings, in this example we have: 3 appears before 1 and 2, 4 appears before 1 and 2, and 2 appears before 1: a total of 5 inversions.

The maximum possible number of inversions is N(N-1)/2 and the minimum possible is 0. The number of inversions is normalized to lie between -1 and 1 by subtracting N(N-1)/4 and dividing by N(N-1)/4. If N=4, as in this case, we get 2/3.

The result is the Spearman rank correlation. Like the dimensionless covariance, it ranges from -1 to 1 with negative scores indicating random variables that move opposite one another while positive scores tend to indicate random variables that move together.

An alternative way to define Spearman rank correlation is to compute the dimensionless covariance of the rankings, it turns out that this is equivalent to the above definition.

Another closely related measure is the Kendall rank correlation coefficient tau ("Kendall's tau"). As before, we take pairs (X, Y) and rank them. Then we count the number of times that the first variable is greater than the second variable (in rank). This count yields a number between 0

and N which we subtract from N/2 and normalize by dividing by N/2 to produce a final result in the range of −1 to 1 as for the other two measures. The result is Kendall's tau. Using the example with four observations above, the only pair with the first coordinate's rank larger than the second's is the pair (3, 2), so we subtract 1 from 4/2 and divide by 4/2. This tells us that Kendall's tau for this data is 0.5.

There is an equivalent way to think of Kendall's tau that gives it a nice intuitive interpretation. Select two pairs of observations (X_1, Y_1) and (X_2, Y_2) from the jointly distributed population. Call this pair concordant if either $X_1 < X_2$ AND $Y_1 < Y_2$ or $X_1 > X_2$ AND $Y_1 > Y_2$; otherwise call the pair discordant (if ties are permitted this definition needs to be adjusted). Let P(C) and P(D) be the probabilities of concordant and discordant observations, respectively. Then Kendall's tau is P(C) – P(D).

In a scatterplot, Kendall's tau can be thought of as the probability that an observation is in quadrants one or three (concordant) minus the probability that it is in quadrants two or four (discordant).

So, Kendall's tau measures the likelihood that two random variables move in the same direction. It is a statement about probability; it does not try to relate the size of the changes, only their directions. I suspect that when people say that two things are correlated, they generally have something like this relationship in mind.

A third non-parametric measure, which we mention in passing is the Gamma statistic. It is similar to Kendall's tau in that it also is a probability. It is useful when there are many ties (such as you might find in frequency data). It is the probability of concordance minus the probability of discordance divided by one minus the probability of ties. (See for example, Siegel & Castellan [SC])

All three of these non-parametric measures have the nice feature that they are equal to one in the case on concordance and minus one in the case of discordance. Also, they are invariant under increasing transformation, so they are actually properties of the copula.

Kendall's tau (τ) and Spearman's rank statistic (R) cannot differ by very much, as illustrated by the following inequality due to Daniels [see N]:

$$-1 \leq 3\tau - 2R \leq 1$$

What this inequality tells us is that if either Kendall's tau or the Spearman R is near 1 (or −1) then the other must be also, and when one is

near 0 the other cannot be too far away. For example, if the Spearman coefficient is zero then Kendall's tau is at most 1/3 in absolute value.

An alternative development of Kendall's tau, Spearman rank correlation, and dimensionless covariance can be found in Wang's paper on aggregation of risk portfolios [W].

Simulation

One of the more common methods for estimating aggregate distributions is to run a Monte Carlo simulation. Reinsurance may have features such as drop-downs or shared retentions, or any number of other features that make analytic modeling practically impossible. In these cases, Monte Carlo simulation may be the only available tool.

Determining how to incorporate correlation information into a simulation is an important problem. It is quite common to use the normal copula in these cases, as we have seen earlier there are some pitfalls associated with this.

Some canned simulation packages can induce a given Spearman rank correlation. One possible method for simulating a given rank correlation structure is described below.

Parameter uncertainty is commonly dealt with through use of a simulation. The incorporation of this uncertainty can produce some unexpected results. Let X and Y be two independent identically distributed random variables that are both normal with variance 1 and mean θ, where θ is either 1 or 2, but is not known which. Suppose that we run a simulation and assign θ the value 1 half of the time and the value 2 the other half of the time, using the same value of θ for both X and Y. We then compute the correlation of X and Y across the simulation. We will show a positive correlation even though X and Y are independent! The scatterplot of the result will look similar to plot number III. (The scatterplot of the two lines of insurance with an urban cluster and a rural cluster.)

In the previous example, θ was constant in each scenario. Frequently, parameter uncertainty arises from uncertainty about the parameter for each individual. This uncertainty can manifest also itself as observed correlation. Meyers [M] has written extensively about this topic.

Even if it is believed that the correlation is known, it may be best to use Kendall's tau or rank correlation anyway. There are several reasons for this. In the case of binormal random variables, there is a formula for converting from dimensionless covariance to rank correlation and vice versa.

So, in at least one important case they are almost the same.[6] Secondly, the confidence intervals that arise in the calculation of dimensionless covariance from samples are generally very wide. If one uses rank correlation, it is possible to achieve the specified rank correlation when performing a Monte Carlo simulation, a goal which in general in unachievable in the case of dimensionless covariance as we have seen.

One possible method to generate random variables with a given rank correlation in a simulation is to do the following:

Assume that the variables are X and Y and that a method exists for generating instances of each of these variables. We will also assume that a simulation with N iterations is to be run.

Step 1: Generate N instances of X and Y. Sort each of these from smallest to largest.

Step 2: Notice that paired this way (sorted) they are fully concordant so the rank correlation is one.

Step 3: (Loop) While the rank correlation is larger than desired, interchange two of the Y's moving a smaller one up and a larger one down.

The choice of pair to swap in step 3 can be tailored to make the process appear to have more or less tail dependence as the modeler desires.

An Example of an Aggregate Stop Loss

To illustrate some of these concepts, let us examine a simplified example of an aggregate stop loss. Our model consists of 100 risks, each of which has a 5% chance of suffering a loss of (exactly) 1. For simplicity, we will assume a zero percent discount rate.

We will divide these losses into three layers: a primary layer that covers losses in aggregate from 0 up to 6, a "working layer" that covers losses from 7 up to 10 and a catastrophe (CAT) layer that covers losses from 11 to 100.

The expected loss to the entire program is 5 (since there are 100 risks each with a 5% chance of loss). Of course, this expectation does not mean

[6] This is the classic case for dimensionless covariance where everything works as expected.

that we expect no loss in the working layer. If the losses were independent Bernoulli trials, the probability of 7 or more losses would be about 23.4% and the probability of 11 or more losses would be less than 0.5%.

What happens if the losses are not independent? The answer is that things can be very different. How they are different depends on the joint distribution function of these 100 variables. We will look at three different models.

First, is it possible for the pairwise correlation to be negative for each pair of insureds? At first blush, ones intuition might lead one to think that you cannot have 100 things which are all pairwise negatively correlated. Positive correlation intuitively means that when one variable goes up the other tends to also, so negative correlation should mean that they move in opposite directions. If we have 100 such variables, then when one of them goes up we expect the other 99 go down ... that means that most of them are moving together! As compelling as this argument is, it is wrong. To see this fallacy, consider our example of 100 risks each with a 5% chance of loss where we are given that the total number of losses is exactly 5. Now, knowledge that one given risk had a loss decreases the probability that any other risk had a loss (from 5 in 100 to 4 in 99). Similarly, knowledge that a given risk did not have a loss increases every other risk's chance of loss (from 5 in 100 to 5 in 99). This information shows pairwise negative correlation. Of course, in this case we could have done a simple calculation to explicitly compute it also (it is equal to $-1/95$).

When we look at the effect of this particular correlation structure on our excess losses we see that it results in no excess losses at all! Since there are always exactly five losses, all of the loss occurs in the primary layer. In this example, negative correlation helps the excess layers and hurts the primary layers. Of course, this example is somewhat unrealistic.

In the second model, we will set the pairwise correlation to a specified value using a normal copula. One way to achieve this pairwise correlation is to generate 101 independent standard normals. One of these variables, say the first one, we will call the "market factor". Each of our one hundred correlated normal random variables will be obtained by taking a multiple of the "market factor" and adding to it one of the other standard normals. The resulting variable will be normal (the sum of normals is normal), but it will require rescaling to be standard (variance 1). If the multipliers are selected appropriately, any specified non-negative correlation can be achieved in this manner.

Here is a simple example. Suppose that we wish to achieve a pairwise correlation of 36%. We create standard normals which are correlated 60%

with the market factor (60% being the square root of 36%). If we call the market factor M, the independent normals X_i, and the correlated normals Y_i, we would set:

$$Y_i = 0.8X_i + 0.6M$$

The Y_i have mean 0 (because both X_i and M do) and variance 1 (because $0.8^2 + 0.6^2 = 1^2$). An easy calculation left to the reader shows that the correlation between distinct Y_i and Y_j is indeed 36%.

These normals are then used to create a copula (i.e. the inverse transformation is used to obtain 100 uniformly distributed (0,1) random variables) the components of which, in turn, are used to determine if the corresponding risk has a loss. If the uniform random variable is less than 0.05, then there is a loss on that risk.

The overall expected losses remain 5, however the sharing of these losses amongst the layers depends heavily on the correlation (however it is measured). The higher the correlation, the more expected loss there is above the primary layer.

Strangely, beyond a certain point additional correlation stops hurting the working layer and actually begins to lower its expected losses. This reduction happens when the additional expected loss coming in from the primary layer is exceeded by the expected loss passed up to the CAT layer. Global reinsurers have observed similar phenomenon when dealing with currencies that have extreme inflation.

In the third model, we will consider the following method for generating losses with a specified pairwise correlation that includes a catastrophe component. Select 101 standard independent normals as before and generate uniform random variables as in the prior example. Now generate one more independent standard normal. If it is smaller than some specified amount, set all of the one hundred risks equal to the first one. It is easy to do the parameter selection so that a given pairwise correlation is exactly achieved.

This model has the exact same correlation as the prior model, but it is vastly more dangerous for the CAT layer and safer for the primary (again the working layer might either benefit or lose). Effectively, our 100 separate risks get replaced by one giant risk a certain percentage of the time. When this giant risk has a loss, the CAT layer is totally wiped out. This is a form of common shock model.

Spectacularly large losses to the CAT layer will almost never be observed in the prior model because of the tail independence of the normal

copula, but they will be observed with this modification. Certain risks (in particular credit losses) may be better modeled by something with a very fat tail, such as a mixture of this last model with the prior one.

Conclusions

Determining how the correlation of risks changes the aggregate loss distribution is a very difficult problem. One number cannot describe the relationship between two jointly distributed random variables. It is possible to make two models with identical correlation coefficients that exhibit wildly different excess aggregate losses. The difference lies in the copula.

The copula (which captures the structure of joint distribution) and the marginals (which capture the individual random variable) can be used to specify pairs (or more generally tuples) of jointly distributed random variables. Casualty actuaries are beginning to use these tools to understand the correlation structure of insurance loss random variables. This use will no doubt increase in the future.

Because of its relationship to the second moment of the aggregate loss distribution, Pearson product moment correlation (which we have referred to as dimensionless covariance) is a fundamental quantity. However, it lacks some of the nice features of the non-parametric measures, Kendall's tau and Spearman rank correlation.

The non-parametric measures may be somewhat more intuitive and may better capture what management is thinking when they say that two things are correlated.

References:

[B] Brehm, P.; "Correlation and the Aggregation of Unpaid Loss Distributions"

[CS] Clark, C. and Schkade, L.; "Statistical Methods for Business Decisions", South-western Publishing Co., 1969

[ELM] Embrechts, P.; Lindskog, F.; McNeil, A.; "Modeling Dependence with Copulas and Applications to Risk Management"

[EMS] Embrechts, P.; McNeil, A.; and Straumann, D; "Correlation and Dependence in risk Management: Properties and Pitfalls"

[F] Feller, W.; "An Introduction to Probability Theory and Its Applications", Wiley, 1968

[M] Meyers, G.; "Estimating Between Line Correlations Generated by Parameter Uncertainty"

[MS] Mango, D. and Sandor, J.; "Dependence Models and the Portfolio Effect"

[N] Nelson, Roger B.; An Introduction to Copulas, Springer, 1999

[SC] Siegel, S. and Castellan, N. J.; Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Company, 1988.

[V] Venter, G.; "Tails of Copulas"

[W] Wang, S.; "Aggregation of Correlated Risk Portfolios: Models and Algorithms", 1999 PCAS LXXXV

Estimating the Parameter Risk of a Loss Ratio Distribution

Charles E. Van Kampen, FCAS, MAAA

Estimating the Parameter Risk of a Loss Ratio Distribution

Chuck Van Kampen, FCAS, MAAA
American Agricultural Insurance Company

Abstract
Actuaries commonly build statistical models to predict future experience. To do this a model must be chosen, and parameters for that model must be calculated and selected. This paper assumes that the correct model has been chosen, but looks at the risk taken by assuming that the selected parameters accurately represent the true underlying distribution. A bootstrapping methodology is used to estimate the parameter risk associated with a loss ratio distribution. The results provide an estimate of the parameter risk of the ground-up loss ratio and for excess loss ratio layers commonly known as aggregate stop loss contracts. The paper shows that the impact of parameter risk on expected losses can be significant especially for aggregate stop loss contracts.

Acknowledgements
I would like to thank Janet Katz and Manalur Sandilya for their help in writing this paper. Janet offered the original encouragement to write the paper, and offered numerous suggestions that kept me moving forward. Manalur offered valuable suggestions on how to present the ideas and procedures more clearly.

Estimating the Parameter Risk of a Loss Ratio Distribution

Introduction

This paper presents a methodology for estimating the parameter risk of a loss ratio distribution. Estimates of parameter risk will be calculated for the ground-up loss ratios and for excess loss ratio layers. The parameter risk estimate is calculated when determining the expected loss of the business being priced.

The basic idea of estimating parameter risk is similar to Hayne [1], and to Meyers and Shenker [2], but instead of a theoretical approach a bootstrapping technique is used. Also different is that the work mentioned above concentrates on the collective risk model with frequency and severity distributions, but this paper uses loss ratios directly. Loss ratios were used because this is frequently all that is available for pricing aggregate stop loss covers, and often used when pricing primary business.

The key idea of the paper is that many different sets of parameters could have produced the actual data, and it is impossible to know which set of parameters produces a model that accurately represents the actual underlying distribution. The methodology presented determines sets of lognormal model parameters that could have produced the given loss ratios, and the relative probability of each of these sets of parameters. These parameters sets and their relative probabilities are then used to determine a ground-up expected loss ratio and to price possible aggregate stop loss reinsurance layers.

The paper will show that an actuary that does not take parameter risk into account runs the risk of underestimating the expected loss. The impact of parameter risk on expected losses can be significant especially for aggregate stop loss contracts.

Assumptions Underlying the Analysis

It was assumed that a simple loss ratio distribution would adequately represent the true underlying exposure of the ceding insurance company. A one in a hundred year catastrophe is likely not included in the data available, but the exposure for this type of loss is still present. The loss experience for

these types of exposures should be removed from the total loss experience, and the expected loss for these exposures should be calculated separately.

It was also assumed that the loss ratios given produce the correct prospective estimate of the ultimate loss, and that there was no risk from reserving or pricing assumptions. This simplification allows the paper to concentrate on the parameter risk of the loss ratio distribution.

The lognormal distribution was used because it is a flexible distribution, and much of the math has been programmed directly into most spreadsheet software. This methodology can be completed using most probability distributions.

It should also be noted that the selected distribution should be a good approximation of the actual prospective loss ratio probability distribution. If there is not a reasonably good fit then this methodology will not provide appropriate results. All models should be checked against the actual data to ensure they are a reasonable representation of that data.

The Prospective Loss Ratio Distribution

Exhibit 1 provides a ten-year prospective loss ratio distribution and the mean, standard deviation and skewness of that distribution. It also provides the logs of the loss ratios, and the mean and standard deviation of the logs of the loss ratios. The mean and standard deviation of the logs were assumed to be the best-fit parameters Mu and Sigma.

It is a good idea to calculate a reasonable pair of best-fit parameters. This can be done using several well-documented methods [3]. The best-fit parameters will act as the starting point for determining viable parameter sets.

Also presented in Exhibit 1 is the actual experience of an aggregate stop loss layer, and the expected loss for the aggregate stop loss layer using the fitted parameters. The Expected Loss on Line is also given. This is the expected loss of the layer divided by the maximum loss of the layer.

Determining Viable Parameter Sets

The intent of the methodology is to find sets of parameters for the lognormal distribution that could have produced the prospective loss ratio distribution, and to determine the relative probability for each set of parameters. These parameter sets and their relative probabilities are then used to directly price the cover.

The methodology first needs to determine what sets of lognormal parameters could have possibly produced the given data. This is done using an excel macro stepping through parameter ranges. The macro methodically steps through ranges around the best-fit Mu and Sigma parameters creating possible parameter sets.

The following procedure was used to determine viable parameter sets.

1.) A Possible Mu and Sigma parameter set is determined from the excel macro.
2.) Ten years of data were used for this analysis; so 10,000 ten-year blocks of loss ratios are simulated using the lognormal distribution with the possible parameter set.
3.) For each ten-year block the simulated mean, standard deviation and skewness is compared to the actual ten-year prospective loss ratio mean, standard deviation and skewness. If it is close then that ten-year block is marked, and the parameter set is considered to be viable.
4.) The number of simulated ten-year blocks that were marked as having a mean, standard deviation and skewness close to the actual prospective loss ratio experience are tallied and recorded along with the parameter set Mu and Sigma values.
5.) The Excel macro moves to the next possible Mu and Sigma parameter set in step 1 until it has stepped through the entire range of possible Mu and Sigma parameters.

The size of the parameter ranges, the step sizes through the ranges and the definition of close are discussed in the Comments on Determining Viable Parameter Sets section of the paper.

A sample of steps 2 through 4 above can be seen in Exhibit 2. The parameter set used for the simulations in Exhibit 2 was the best-fit parameters with a Mu parameter of -0.45 and a Sigma parameter of 0.11. In

the bottom section of Exhibit 2 each row is a simulated 10-year block with the mean, standard deviation and skewness of the simulated 10-year block given. These are compared to the ranges determined and labeled above as "Min Target Range" and "Max Target Range". If the simulated mean, standard deviation and skewness are all within the target ranges determined it shows up as a 1 in the "Frequency" column at the far right. Just one row of simulated loss ratios in Exhibit 2 meets all three criteria.

A sample of the outcome from the entire process can be seen in Exhibit 3. The parameter set test number of 1600 means that -0.45 and 0.11 were the 1600th set of parameters tested. The frequency of 1600 is 117. This is how many times out of the 10,000 simulations that the simulated ten-year loss ratio mean, standard deviation and skewness were close to the actual prospective loss ratio mean, standard deviation and skewness. Finally the relative frequency is given. For parameter set 1600 the relative probability is 2.985%. This is calculated by taking the frequency of parameter set 1600 (117) divided by the total frequency of all parameter sets (not given but is 3,920). The parameter sets with frequencies greater than zero are sorted out and used for pricing.

In Exhibit 3 it can be seen that the Mu parameter was -0.45 and Sigma varies from 0.0055 to 0.4345. Some insignificant entries were removed so that the exhibit is easier to read. There are 78 total steps through the Sigma range. The Mu parameter also steps through a range, but this is not shown in Exhibit 3 where it is just –0.45. The range that the Mu parameter steps through is bounded by –1.549 to 0.226, and there are 50 steps through the Mu parameter range.

In the bottom section of Exhibit 4 the parameter sets have been sorted by relative probabilities, and the sets with the largest relative probabilities are presented. The best-fit parameters of -0.45 under Mu and 0.11 under Sigma were the third most likely set of parameters to simulate close 10-year blocks. Using 10,000 simulations did not ensure that the best-fit pair was the most likely, but it does produce a relative probability that is close to what it should be. More simulations could be used to increase accuracy, but there are trade-offs between the processing time and the accuracy.

The important thing to note is not which parameter set is most likely, but how many possible parameter sets had large relative probabilities. In total

there are 364 parameter sets out of 3950 parameter sets tested that simulated at least one ten-year block close to the actual data provided.

Graphs of the parameter sets and their relative probabilities are given in Figures 1and 2. Each of the graphs shows the relative probabilities of the parameter sets from different perspectives. Figure 1 shows the distribution from the side, and Figure 2 is a top view. These graphs must be read carefully because the scale for the Mu and Sigma are determined by the step sizes selected. Note that the scales were selected so that all of the parameter sets greater than zero could be seen.

In Figures 1 and 2 it can be seen that the Sigma parameter has a significant amount of skew. It should also be pointed out that Mu and Sigma are related. Given a certain Mu only Sigma parameters within a certain range will provide simulated 10-year blocks that are close to the given mean, standard deviation and skewness, and within that Sigma range some Sigma parameters are much more likely than others.

Calculating the Estimate of Parameter Risk

Calculating the expected loss using the viable parameter sets is a reasonably straightforward task. The lognormal expected value formula is used for each viable parameter set to come up with the expected loss for the parameter set. Each parameter set expected loss is then weighted together using the relative probabilities determined earlier in the process.

Exhibit 4 shows the expected loss for the best-fit pricing methodology and the parameter set methodology. The top block labeled "Fitted Original Distribution" gives the expected loss for the best-fit pricing, and the second block labeled "Parameter Set Distributions" gives the summary of the expected loss for the parameter set pricing. The difference between the expected losses of the two methods is labeled as "Difference Parameter Set to Fitted". This is calculated as (Parameter Set Expected Loss/Fitted Expected Loss) – 1. The weighted expected loss for the most likely parameter sets is shown in the lower section of the exhibit. The first column of calculated expected losses gives the ground-up expected loss for the distribution, and the second column gives the expected loss of the aggregate stop loss layer given in Exhibit 1. Each column after that gives the expected loss for the given loss ratio range.

In Exhibit 4 it can be seen that for a primary insurer the ground up best-fit expected loss ratio is 64.1% (Exp Loss Ratio = Exp(Mu+(Sigma^2)/2)), but the parameter set expected loss is 64.5%. If the company's permissible loss ratio is 63.0% than the best-fit indicated rate change is 1.75%, but the parameter set indicated rate change is 2.38%. If the margin for the risk and profit of this business is 5.0%, than the difference between the two indications is 12.38% of the margin including the lower rate increase ((2.38%-1.75%)/(5% of 101.75%)).

The aggregate stop loss layer given in Exhibit 1 is the second column in Exhibit 4. It has an attachment point of 72.5% and a layer of 2.5%. It can be seen in Exhibit 4 that there is a 45.85% difference or load in expected loss for a 2.5% excess 72.5% aggregate stop loss layer. The additional columns show how the difference in expected losses changes by layer, and how it increases significantly as the expected loss on line goes down.

Note that the difference in expected losses is negative for the low layers up to 65%. It appears that the parameter set methodology is shifting the probability distribution from the lower loss ratios to the higher loss ratios where aggregate stop loss contracts are usually purchased.

The difference in expected losses is the key finding of the paper. It is the estimated parameter risk for the prospective loss ratio distribution given. If the actuary does not calculate and include an estimate of parameter risk in the price then the company may not be charging enough for the business.

Sensitivity Analysis

Several alternative situations were looked at using the parameter set methodology. A higher mean, a larger standard deviation, a larger skewness and decreasing the number of years of data were investigated. Loads for primary expected losses and expected loss by layer are both considered.

The other alternatives are presented in Exhibits 5 through 8. Exhibit 5 increases the expected mean of the actual data (-0.36 0.11 0.5). Exhibit 6 increases the standard deviation of the actual data (-0.45, 0.2 0.5). Exhibit 7 increases the skewness of the actual data (-0.45, 0.11 1.5). Exhibit 8 assumes that only five years worth of data were available for pricing but that the mean, standard deviation and skewness remained the same (-0.45, 0.11 0.5 5Yr).

If the loads by layer are looked at by expected loss on line (expected loss for the layer/maximum loss of the layer) they can be compared to distributions with different means and standard deviations. In Exhibits 4 through 8 the expected loss on line is given for both the best-fit pricing and the parameter set pricing. A graph comparing the loads by expected loss on line for the parameter set pricing is given in Figure 3.

Parameter set relative probability graphs for the four additional distributions are given in Figures 4 through 11. The same views are presented for each distribution. Please note that these must be looked at carefully because the scales are not consistent. Looking at the relative probabilities should help compare between the four distributions given. Using only five years of data required that the Sigma parameter range be extended. This can be seen in Figures 10 and 11.

The actual data underlying each of these additional distributions is given in Exhibits 9 through 12. The format of these exhibits is the same as Exhibit 1.

Increasing the Mean

For both the primary and layered loss ratios the load does not vary substantially when the mean is increased (Exhibit 5 and Figure 3 with -0.36 0.11 0.5). This would intuitively make sense. The shape of the distribution has not changed, but is just shifted upwards.

Increasing the Standard Deviation

For the alternative with the higher standard deviation (Exhibit 6 with -0.45 0.2 0.5) the difference in indicated ground-up rate changes increased significantly. The ground-up fitted loss ratio went from 65.1% for the best-fit pricing to 66.5% for the parameter set pricing. The difference between rate change indications based on the same 63% permissible loss ratio rose to 2.23%, or 43.16% of a 5% margin. This implies that having a block of information with loss ratios that are less stable has a substantial impact on the ground-up parameter risk.

The load by layer is somewhat higher for the distribution with the higher standard deviation (Figure 3 -0.45 0.2 0.5). One possible explanation for this is that increasing the standard deviation shifts the parameter set distribution more than the best-fit distribution.

Increasing the Skewness

Increasing the skewness (Exhibit 7 with 0.45 0.11 1.5) does not seem to have a large impact on the ground-up loss ratios. The difference in rate change indications rose to 0.79%, or 15.53% of a 5% margin.

Looking at Figure 3 (0.45 .11 1.5) the load for excess layers seems to track closely with the loads for the higher standard deviation. It appears that increasing the skewness has an impact similar as increasing the standard deviation of the actual data on the excess layers.

Five Years of Data Available

Only having five years of data (Exhibit 8 with –0.45, 0.11 0.5 5Yr) has a significant impact on the indicated primary loss ratio parameter risk load. The ground-up loss ratio went from 64.1% for the best-fit pricing to 65.7% for the parameter set pricing. The difference in rate change indications rose to 2.54% or 49.93% of a 5% margin.

Looking at the five-year distribution (Figure 3 –0.45 0.11 0.5 5Yr) it can be seen that for every expected loss on line that the load by layer is significantly higher for the five-year distribution. It appears that having fewer years of data has a much larger impact on loads for the higher layers than increasing the standard deviation or skewness.

The number of years of data is a key input into the model. The actuary should attempt to find additional years of data, or find other alternative sources of data that are reasonably consistent with the available data. Integrating this additional data into the analysis should bring the parameter risk down.

Comments on Determining Viable Parameter Sets

It is noted that the above process does contain a fair amount of judgment. The distance of the steps in the parameter ranges, the size of the parameter ranges and the definition of close should be discussed. Also the use of mean, standard deviation and skewness as the criteria to judge if a parameter set is close should be considered.

Step Size

The size of the steps through the parameter ranges should be considered. Running the Excel macro several times using increasingly smaller step sizes each time can help determine an appropriate step size. When the results don't change substantially the step size is small enough to be a reasonable estimate of the infinite number of possible parameter sets. The macro used to calculate the information in Exhibit 4 took around 8 hours to run. When the step sizes were cut in half and the macro was rerun it took 36 hours to complete the macro. The results are presented in Exhibit 13. The expected primary loss ratio stayed at 64.5%, and there was an increase of only 0.001% when the loss ratios are rounded to hundred-thousandths. The expected loss to the layer from 72.5% to 75.0% went from 0.342% to 0.345%, or an increase of 0.703%. The change in results does not seem to justify the increase in processing time.

Size of Ranges

The parameter set ranges that the Excel macro loops through should be wide enough so that at the edges of the ranges very few simulated ten-year blocks have a mean, standard deviation and a skewness close to the actual mean, standard deviation and skewness.

Close to Original Distribution

The intent of close should also be considered. The definition of close selected was based on practical considerations. How close does the simulated ten-year block need to be to the actual mean and standard deviation in order to be comfortable while pricing the underlying cover. Skewness was included to make sure that the distribution had the correct skew around the mean. These considerations have to be balanced against the frequency needed to minimize the impact that random simulations could have on the outcome.

To be consistent the final definition of close was that when using the best-fit parameters about 21%-22% of the simulated means, standard deviations and skews were within a band centered at the actual mean, standard deviations and skewness of the actual data. For the distribution in Exhibit 2 this was a mean centered at 64.1% with a band from 63.52% to 64.77%. The standard deviation was centered at 0.0712 with a band from 0.0662 to 0.0762. The

skewness was centered at 0.5 with a band from 0.29 to 0.71. This translates into roughly 100 simulated viable 10-year blocks when using the best-fit parameters (10,000 simulations x 21% x 21% x 21%).

Criteria Used

The mean, standard deviation and skewness were used as the criteria to determine if the parameter set is viable. Other measures could have been used in place of or in addition to these. The mean, standard deviation and skewness were used because they are well known, are simple to work with, and capture the basic characteristics of the underlying loss ratio distribution. Other measures should be considered.

Conclusion

Viable parameter sets and their relative probabilities were determined, and then used to directly calculate the expected primary loss ratios and the expected loss by layer. The concept is that any one of the viable parameter sets could have produced the actual loss ratio experience of the ceding company. By comparing the expected losses of the parameter sets to the expected losses of the original best-fit an estimate of the parameter risk of using just one set of parameters can be determined.

The process requires a fair amount of judgment, but actuarial pricing cannot be completed without some level of judgment. The assumptions that need to be made are relatively straightforward and can easily be changed to measure the impact.

It should also be mentioned that this methodology only estimates the parameter risk of the loss experience that is present in the prospective loss ratios. Actuaries hope that modeling will account for some of the potential variation that could occur in the experience. There are exposures that did not have loss experience within the data set, and that are not anticipated nor reflected in the adjusted loss ratios. This risk is still present and not measured directly by the methodology. The catastrophe and other risk that was removed in the beginning still need to be taken into account. Expected losses for these exposures need to be determined and included in the overall expected loss.

In addition the original data set likely has process risk included in the experience. Even when using the best-fit parameters to simulate 10-year blocks of loss ratios only a small number of the 10-year blocks had a mean, standard deviations and skewness close to the actual data. This would imply that the process risk could be substantial. This paper assumes that the given data provides a good representation of the true underlying distribution.

This methodology could be adapted to most statistical distributions. It therefore could apply to a wide variety of situation where statistical distributions are used. Property and casualty primary loss ratios and an aggregate stop loss reinsurance layer were looked at in this paper, but there is no reason it can't be used in other situations to estimate parameter risk.

References

[1] Hayne, R.M., "Modeling Parameter Uncertainty in Cash Flow Projections", *Casualty Actuarial Society Forum*, Summer 1999, pp. 133-151.

[2] Meyers, G.G., Schenker, N., "Parameter Uncertainty in the Collective Risk Model", *Proceedings of the Casualty Actuarial Society*, LXX, 1983, pp.111-143.

[3] Hogg, R. and Klugman, S. (1984). *Loss Distributions.* John Wiley & Sons, Inc.

Loss Ratio Distribution

Exhibit 1

	58.4%	
	Actual	Actual
	LR	Ln(LR)
1	58.4%	(0.5376)
2	64.5%	(0.4388)
3	67.3%	(0.3953)
4	52.6%	(0.6415)
5	58.4%	(0.5376)
6	64.5%	(0.4388)
7	78.3%	(0.2440)
8	70.6%	(0.3488)
9	62.0%	(0.4786)
10	64.5%	(0.4388)
Average	64.1%	(0.4500)
Stdev	0.0712	0.1100
Skew	0.5000	
Expected		64.1%

Min	72.5%
Max	75.0%
	Loss
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	2.5%
	0.0%
	0.0%
	0.0%
Average	0.250%
Stdev	0.0079
Expected Loss On Line	10.0%
Fitted Expected Loss	0.235%
Fitted Expected Loss On Line	9.389%

Simulating Loss Distributions Exhibit 2

		Sample Values	Expected Value of X	LN(X)	
Distributi	LnN	64.1%	64.1%	(0.4500)	Mean
Param 1	(0.4500)	0.0712		0.1100	Standard Devation
Param 2	0.1100	0.5000			Skewness
Prob XS			0.625%	0.6352	Min Target Range Mean
Count	4000			0.6477	Max Target Range Mean
			0.500%	0.0662	Min Target Range Std Dev
				0.0762	Max Target Range Std Dev
			21.000%	0.2900	Min Target Range Skew
				0.7100	Max Target Range Skew

117

Simulated Year										Mean	StdDev	Skewness	Mu	Std Dev	Skew	Frequency
1	2	3	4	5	6	7	8	9	10	**0.6411**			(0.4500)	0.1100	0.50000	
													22.00%	21.97%	21.36%	1.17%
59.7%	67.5%	59.7%	72.8%	67.9%	70.0%	63.5%	71.6%	70.0%	70.4%	0.6730	0.047	(0.8060)	0	0	0	0
67.9%	59.0%	61.1%	62.1%	63.7%	55.6%	66.9%	71.4%	66.8%	64.6%	0.6390	0.046	(0.2625)	1	0	0	0
72.3%	58.4%	49.1%	61.0%	51.8%	75.1%	53.6%	67.2%	79.4%	59.3%	0.6270	0.104	0.3481	0	0	1	0
67.0%	61.5%	70.3%	57.2%	75.0%	61.8%	72.9%	61.6%	54.3%	66.2%	0.6478	0.067	0.0551	0	1	0	0
80.2%	77.3%	61.5%	66.8%	72.5%	72.7%	69.9%	61.1%	65.6%	78.6%	0.7063	0.068	(0.0387)	0	1	0	0
65.2%	67.4%	62.3%	60.4%	62.5%	58.0%	64.9%	62.7%	60.1%	62.2%	0.6259	0.027	0.1428	0	0	0	0
64.5%	49.1%	56.5%	54.4%	70.3%	71.4%	60.8%	57.7%	63.1%	55.7%	0.6033	0.071	0.2402	0	1	0	0
63.7%	66.3%	70.7%	68.0%	58.3%	73.6%	60.6%	59.3%	62.8%	68.8%	0.6521	0.051	0.1746	0	0	0	0
69.1%	63.9%	65.5%	70.5%	62.3%	76.7%	66.0%	59.7%	68.7%	50.9%	0.6532	0.069	(0.6321)	0	1	0	0
60.4%	50.3%	54.8%	73.9%	52.9%	56.0%	65.0%	87.8%	55.8%	55.7%	0.6124	0.115	1.6355	0	0	0	0
67.1%	63.0%	66.8%	55.3%	68.1%	72.6%	60.5%	55.6%	53.6%	53.0%	0.6157	0.070	0.1487	0	1	0	0
69.3%	54.5%	58.6%	66.4%	52.1%	67.9%	70.5%	68.6%	59.0%	81.9%	0.6487	0.089	0.3172	0	0	1	0
60.8%	69.1%	69.4%	65.3%	69.4%	69.0%	59.4%	73.2%	61.7%	64.7%	0.6623	0.045	(0.1600)	0	0	0	0
68.7%	63.4%	60.0%	64.2%	71.6%	63.0%	71.4%	63.4%	63.8%	57.2%	0.6466	0.046	0.2322	1	0	0	0
60.6%	57.7%	57.6%	70.9%	55.7%	74.9%	66.6%	63.9%	61.1%	73.9%	0.6428	0.070	0.4429	1	1	1	0
56.9%	61.4%	58.2%	69.7%	72.7%	61.4%	66.9%	70.1%	61.8%	76.5%	0.6556	0.068	0.2973	0	0	1	0
56.8%	62.2%	59.7%	67.5%	64.3%	69.2%	59.0%	70.7%	62.1%	61.6%	0.6331	0.046	0.3840	0	0	1	0
58.8%	54.2%	50.6%	62.1%	61.3%	79.8%	69.7%	57.3%	58.2%	53.0%	0.6050	0.086	1.3500	0	0	0	0
62.7%	66.4%	69.7%	73.3%	65.1%	72.3%	61.9%	60.0%	70.7%	76.5%	0.6789	0.055	0.0395	0	0	0	0
61.6%	67.9%	66.6%	63.3%	63.0%	65.0%	65.9%	57.9%	71.7%	72.7%	0.6557	0.045	0.1061	0	0	0	0
66.2%	60.3%	61.2%	66.5%	75.5%	55.4%	67.8%	61.8%	50.3%	55.9%	0.6209	0.073	0.2043	0	1	0	0
72.5%	46.4%	69.1%	75.4%	54.4%	78.9%	58.2%	85.2%	73.5%	66.9%	0.6804	0.119	(0.5527)	0	0	0	0
55.5%	59.9%	62.4%	60.7%	65.6%	49.7%	50.5%	62.2%	52.5%	57.3%	0.5763	0.055	(0.2139)	0	0	0	0
63.1%	65.8%	61.7%	59.9%	60.4%	60.4%	64.9%	59.3%	64.9%	56.9%	0.6175	0.029	(0.0099)	0	0	0	0
71.4%	63.7%	65.3%	59.2%	74.3%	52.1%	72.7%	72.2%	65.1%	59.9%	0.6558	0.072	(0.5295)	0	1	0	0
61.4%	72.7%	65.1%	63.0%	61.0%	68.0%	74.5%	56.5%	64.1%	59.5%	0.6458	0.057	0.6148	1	0	1	0
54.9%	59.8%	70.8%	62.4%	68.7%	59.9%	58.7%	63.0%	63.9%	64.7%	0.6268	0.047	0.2350	0	0	0	0
59.4%	68.7%	71.9%	66.1%	71.2%	63.8%	56.5%	61.9%	69.5%	65.9%	0.6550	0.051	(0.4868)	0	0	0	0
63.8%	64.0%	59.5%	56.8%	54.0%	80.4%	54.1%	54.3%	58.9%	59.1%	0.6051	0.079	2.0171	0	0	0	0
64.5%	69.4%	73.6%	72.0%	80.4%	71.9%	56.3%	60.7%	57.3%	68.9%	0.6750	0.077	(0.0749)	0	0	0	0
66.2%	54.7%	62.0%	68.5%	64.5%	66.9%	55.7%	52.0%	60.5%	54.5%	0.6055	0.060	(0.0986)	0	0	0	0
80.6%	59.7%	59.9%	59.4%	62.7%	57.7%	70.1%	51.4%	59.3%	55.7%	0.6164	0.082	1.5169	0	0	0	0
62.8%	54.7%	62.3%	58.9%	71.4%	65.1%	70.9%	64.7%	68.0%	56.8%	0.6356	0.057	(0.1001)	1	0	0	0
57.2%	69.7%	63.9%	71.4%	66.7%	61.7%	52.5%	74.0%	66.5%	57.3%	0.6408	0.069	(0.2704)	1	1	0	0
75.2%	61.1%	63.2%	59.7%	53.4%	68.0%	57.1%	60.3%	69.0%	65.1%	0.6322	0.063	0.4329	0	0	1	0

Sample of Simulated Outcomes Exhibit 3

Parameter Set Test Number	Mu	Sigma	Frequency	Relative Probability
1581	-0.45	0.0055	-	0.000%
1582	-0.45	0.0110	-	0.000%
....	....	....	....	....
1592	-0.45	0.0660	10	0.255%
1593	-0.45	0.0715	22	0.561%
1594	-0.45	0.0770	43	1.097%
1595	-0.45	0.0825	82	2.092%
1596	-0.45	0.0880	99	2.526%
1597	-0.45	0.0935	103	2.628%
1598	-0.45	0.0990	121	3.087%
1599	-0.45	0.1045	118	3.010%
1600	-0.45	0.1100	117	2.985%
1601	-0.45	0.1155	102	2.602%
1602	-0.45	0.1210	89	2.270%
1603	-0.45	0.1265	90	2.296%
1604	-0.45	0.1320	79	2.015%
1605	-0.45	0.1375	59	1.505%
1606	-0.45	0.1430	38	0.969%
1607	-0.45	0.1485	49	1.250%
1608	-0.45	0.1540	28	0.714%
1609	-0.45	0.1595	30	0.765%
1610	-0.45	0.1650	19	0.485%
1611	-0.45	0.1705	24	0.612%
1612	-0.45	0.1760	15	0.383%
1613	-0.45	0.1815	13	0.332%
1614	-0.45	0.1870	10	0.255%
1615	-0.45	0.1925	11	0.281%
1616	-0.45	0.1980	6	0.153%
1617	-0.45	0.2035	5	0.128%
1618	-0.45	0.2090	6	0.153%
1619	-0.45	0.2145	5	0.128%
1620	-0.45	0.2200	-	0.000%
1621	-0.45	0.2255	5	0.128%
1622	-0.45	0.2310	2	0.051%
1623	-0.45	0.2365	2	0.051%
....	....	....	....	....
1658	-0.45	0.4290	-	0.000%
1659	-0.45	0.4345	-	0.000%

Fitted Distributions - LnNormal
Base

Exhibit 4

	E(X)	Mu	Sigma	Skew
Fitted	64.1%	(0.4500)	0.11	0.5

Fitted Original Distribution	**Ground-Up**	**Stop Loss Contract**									
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**64.1%**	**0.235%**	49.975%	4.785%	4.105%	2.858%	1.532%	0.629%	0.201%	0.052%	0.011%	0.002%	0.000%
Expected Loss On Line		**9.389%**	99.950%	95.707%	82.091%	57.152%	30.637%	12.572%	4.030%	1.040%	0.223%	0.041%	0.007%

Parameter Set Distributions	**Ground-Up**	**Stop Loss**									
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**64.5%**	**0.342%**	49.860%	4.622%	3.955%	2.850%	1.677%	0.841%	0.388%	0.176%	0.082%	0.039%	0.020%
Expected Loss On Line		**13.693%**	99.719%	92.445%	79.103%	56.994%	33.544%	16.812%	7.754%	3.516%	1.630%	0.788%	0.401%
Difference Parameter Set to Fitted	**0.60%**	**45.85%**	-0.23%	-3.41%	-3.64%	-0.28%	9.49%	33.73%	92.41%	238.17%	632.45%	1838.02%	6058.85%

Mu	Sigma	Freq	Relative Prob											
(0.4500)	0.0990	1.210%	3.087%	1.978%	0.006%	1.543%	0.150%	0.130%	0.089%	0.044%	0.016%	0.004%	0.001%	0.000%	0.000%	0.000%
(0.4500)	0.1045	1.180%	3.010%	1.930%	0.006%	1.505%	0.145%	0.125%	0.087%	0.045%	0.017%	0.005%	0.001%	0.000%	0.000%	0.000%
(0.4500)	0.1100	1.170%	2.985%	1.915%	0.007%	1.492%	0.143%	0.123%	0.085%	0.046%	0.019%	0.006%	0.002%	0.000%	0.000%	0.000%
(0.4500)	0.0935	1.030%	2.628%	1.683%	0.004%	1.314%	0.128%	0.113%	0.077%	0.036%	0.012%	0.003%	0.000%	0.000%	0.000%	0.000%
(0.4500)	0.1155	1.020%	2.602%	1.670%	0.007%	1.300%	0.124%	0.105%	0.074%	0.041%	0.018%	0.006%	0.002%	0.000%	0.000%	0.000%
(0.4500)	0.0880	0.990%	2.526%	1.617%	0.003%	1.263%	0.124%	0.110%	0.074%	0.034%	0.010%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.4500)	0.1265	0.900%	2.296%	1.476%	0.007%	1.147%	0.107%	0.091%	0.065%	0.038%	0.018%	0.007%	0.003%	0.001%	0.000%	0.000%
(0.4500)	0.1210	0.890%	2.270%	1.458%	0.007%	1.134%	0.107%	0.091%	0.064%	0.037%	0.017%	0.006%	0.002%	0.001%	0.000%	0.000%
(0.4172)	0.1100	0.830%	2.117%	1.404%	0.008%	1.058%	0.103%	0.094%	0.072%	0.044%	0.021%	0.008%	0.002%	0.001%	0.000%	0.000%
(0.4500)	0.0825	0.820%	2.092%	1.338%	0.002%	1.046%	0.103%	0.093%	0.062%	0.026%	0.007%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.4172)	0.1045	0.800%	2.041%	1.352%	0.007%	1.020%	0.100%	0.092%	0.070%	0.042%	0.019%	0.006%	0.002%	0.000%	0.000%	0.000%
(0.4500)	0.1320	0.790%	2.015%	1.296%	0.007%	1.006%	0.093%	0.079%	0.056%	0.034%	0.017%	0.007%	0.003%	0.001%	0.000%	0.000%
(0.4839)	0.1155	0.720%	1.837%	1.140%	0.003%	0.917%	0.084%	0.066%	0.042%	0.020%	0.008%	0.002%	0.001%	0.000%	0.000%	0.000%
(0.4839)	0.1210	0.690%	1.760%	1.093%	0.003%	0.878%	0.079%	0.063%	0.040%	0.020%	0.008%	0.003%	0.001%	0.000%	0.000%	0.000%
(0.4172)	0.1210	0.670%	1.709%	1.134%	0.008%	0.854%	0.083%	0.074%	0.057%	0.036%	0.019%	0.008%	0.003%	0.001%	0.000%	0.000%
(0.4839)	0.1100	0.650%	1.658%	1.028%	0.002%	0.828%	0.076%	0.061%	0.037%	0.017%	0.006%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.4172)	0.1320	0.640%	1.633%	1.085%	0.008%	0.816%	0.078%	0.069%	0.053%	0.035%	0.019%	0.009%	0.004%	0.001%	0.000%	0.000%
(0.4172)	0.0990	0.630%	1.607%	1.064%	0.005%	0.804%	0.079%	0.073%	0.056%	0.033%	0.014%	0.004%	0.001%	0.000%	0.000%	0.000%
(0.4172)	0.1155	0.630%	1.607%	1.066%	0.007%	0.803%	0.078%	0.070%	0.054%	0.034%	0.017%	0.007%	0.002%	0.001%	0.000%	0.000%

Fitted Distributions - LnNormal Exhibit 5

Higher Mean

	E(X)	Mu	Sigma	Skew
Fitted	70.2%	(0.3600)	0.11	0.5

Fitted Original Distribution	**Ground-Up**	**Stop Loss Contract**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**70.2%**	**0.769%**	49.998%	4.970%	4.785%	4.183%	3.083%	1.831%	0.867%	0.332%	0.105%	0.028%	0.006%
Expected Loss On Line		**30.775%**	99.996%	99.402%	95.700%	83.667%	61.670%	36.623%	17.349%	6.635%	2.093%	0.558%	0.129%

Parameter Set Distributions	**Ground-Up**	**Stop Loss**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**70.6%**	**0.842%**	49.951%	4.875%	4.615%	4.018%	3.041%	1.942%	1.070%	0.539%	0.263%	0.130%	0.067%
Expected Loss On Line		**33.675%**	99.903%	97.508%	92.292%	80.358%	60.820%	38.850%	21.406%	10.776%	5.263%	2.604%	1.337%
Difference Parameter Set to Fitted	**0.59%**	**9.42%**	-0.09%	-1.91%	-3.56%	-3.96%	-1.38%	6.08%	23.39%	62.41%	151.42%	366.63%	939.02%

Mu	**Sigma**	**Freq**	**Relative Prob**											
-36.0%	11.0%	1.240%	3.273%	0.023	0.025%	1.636%	0.163%	0.157%	0.137%	0.101%	0.060%	0.028%	0.011%	0.003%	0.001%	0.000%
-36.0%	9.9%	1.170%	3.088%	0.022	0.022%	1.544%	0.154%	0.150%	0.133%	0.097%	0.054%	0.023%	0.007%	0.002%	0.000%	0.000%
-36.0%	10.5%	1.140%	3.009%	0.021	0.022%	1.504%	0.150%	0.145%	0.128%	0.094%	0.054%	0.024%	0.009%	0.002%	0.001%	0.000%
-36.0%	12.1%	1.010%	2.666%	0.019	0.022%	1.333%	0.132%	0.125%	0.109%	0.081%	0.050%	0.026%	0.011%	0.004%	0.001%	0.000%
-36.0%	11.6%	0.950%	2.507%	0.018	0.020%	1.254%	0.124%	0.119%	0.103%	0.077%	0.047%	0.023%	0.010%	0.003%	0.001%	0.000%
-36.0%	9.4%	0.830%	2.191%	0.015	0.015%	1.095%	0.109%	0.107%	0.096%	0.070%	0.038%	0.015%	0.004%	0.001%	0.000%	0.000%
-36.0%	12.7%	0.780%	2.059%	0.014	0.017%	1.029%	0.101%	0.096%	0.083%	0.062%	0.039%	0.021%	0.010%	0.004%	0.001%	0.000%
-36.0%	8.8%	0.760%	2.006%	0.014	0.013%	1.003%	0.100%	0.099%	0.089%	0.065%	0.034%	0.012%	0.003%	0.001%	0.000%	0.000%
-32.7%	11.0%	0.720%	1.900%	0.014	0.020%	0.950%	0.095%	0.093%	0.085%	0.069%	0.046%	0.025%	0.011%	0.004%	0.001%	0.000%
-32.7%	9.9%	0.710%	1.874%	0.014	0.019%	0.937%	0.094%	0.092%	0.086%	0.070%	0.045%	0.022%	0.008%	0.003%	0.001%	0.000%
-32.7%	10.5%	0.710%	1.874%	0.014	0.019%	0.937%	0.094%	0.092%	0.085%	0.069%	0.045%	0.023%	0.010%	0.003%	0.001%	0.000%
-32.7%	11.6%	0.710%	1.874%	0.014	0.020%	0.937%	0.093%	0.091%	0.083%	0.067%	0.045%	0.025%	0.012%	0.005%	0.002%	0.000%
-39.4%	11.6%	0.670%	1.768%	0.012	0.010%	0.884%	0.087%	0.081%	0.066%	0.044%	0.024%	0.010%	0.004%	0.001%	0.000%	0.000%
-32.7%	12.1%	0.660%	1.742%	0.013	0.019%	0.871%	0.087%	0.084%	0.076%	0.061%	0.042%	0.024%	0.012%	0.005%	0.002%	0.001%
-39.4%	9.9%	0.650%	1.715%	0.012	0.008%	0.858%	0.085%	0.081%	0.066%	0.043%	0.020%	0.007%	0.002%	0.000%	0.000%	0.000%
-39.4%	12.7%	0.650%	1.715%	0.012	0.010%	0.857%	0.084%	0.077%	0.062%	0.043%	0.025%	0.012%	0.005%	0.002%	0.001%	0.000%
-39.4%	13.8%	0.650%	1.715%	0.012	0.011%	0.857%	0.083%	0.075%	0.061%	0.043%	0.026%	0.014%	0.006%	0.003%	0.001%	0.000%
-36.0%	13.2%	0.610%	1.610%	0.011	0.014%	0.805%	0.079%	0.074%	0.064%	0.048%	0.031%	0.017%	0.008%	0.004%	0.001%	0.000%
-39.4%	12.1%	0.600%	1.584%	0.011	0.009%	0.792%	0.077%	0.071%	0.058%	0.039%	0.022%	0.010%	0.004%	0.001%	0.000%	0.000%

Fitted Distributions - LnNormal
Higher Standard Deviation

Exhibit 6

	E(X)	Mu	Sigma	Skew
Fitted	65.1%	(0.4500)	0.2000	0.5

Fitted Original Distribution	**Ground-Up**	**Stop Loss Contract**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**65.1%**	**0.584%**	49.500%	4.163%	3.483%	2.700%	1.945%	1.309%	0.830%	0.499%	0.287%	0.159%	0.086%
Expected Loss On Line		**23.379%**	98.999%	83.262%	69.653%	54.009%	38.896%	26.176%	16.591%	9.984%	5.748%	3.187%	1.713%
Parameter Set Distributions	**Ground-Up**	**Stop Loss**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**66.5%**	**0.695%**	49.041%	4.013%	3.414%	2.737%	2.081%	1.516%	1.074%	0.749%	0.520%	0.362%	0.254%
Expected Loss On Line		**27.793%**	98.082%	80.258%	68.272%	54.749%	41.610%	30.318%	21.473%	14.977%	10.398%	7.242%	5.087%
Difference Parameter Set to Fitted	**2.23%**	**18.88%**	-0.93%	-3.61%	-1.98%	1.37%	6.98%	15.82%	29.43%	50.02%	80.91%	127.21%	196.99%

Mu	Sigma	Freq	Relative Prob											
-45.0%	19.0%	1.2%	1.665%	0.011	0.009%	0.826%	0.070%	0.059%	0.045%	0.032%	0.021%	0.013%	0.007%	0.004%	0.002%	0.001%
-41.7%	19.0%	1.2%	1.610%	0.011	0.011%	0.801%	0.071%	0.061%	0.049%	0.036%	0.025%	0.016%	0.010%	0.006%	0.003%	0.002%
-41.7%	18.0%	1.2%	1.583%	0.011	0.011%	0.788%	0.071%	0.061%	0.049%	0.035%	0.024%	0.015%	0.008%	0.005%	0.002%	0.001%
-45.0%	20.0%	1.1%	1.542%	0.010	0.009%	0.763%	0.064%	0.054%	0.042%	0.030%	0.020%	0.013%	0.008%	0.004%	0.002%	0.001%
-48.4%	21.0%	1.1%	1.501%	0.009	0.007%	0.739%	0.058%	0.047%	0.036%	0.025%	0.017%	0.010%	0.006%	0.004%	0.002%	0.001%
-48.4%	19.0%	1.0%	1.419%	0.009	0.006%	0.701%	0.057%	0.046%	0.033%	0.023%	0.014%	0.008%	0.004%	0.002%	0.001%	0.001%
-45.0%	22.0%	1.0%	1.392%	0.009	0.009%	0.687%	0.056%	0.047%	0.037%	0.028%	0.020%	0.013%	0.008%	0.005%	0.003%	0.002%
-45.0%	17.0%	1.0%	1.365%	0.009	0.007%	0.679%	0.059%	0.050%	0.037%	0.025%	0.015%	0.009%	0.005%	0.002%	0.001%	0.000%
-45.0%	18.0%	1.0%	1.351%	0.009	0.007%	0.671%	0.058%	0.048%	0.037%	0.025%	0.016%	0.010%	0.005%	0.003%	0.001%	0.001%
-41.7%	21.0%	1.0%	1.338%	0.009	0.010%	0.663%	0.057%	0.050%	0.040%	0.030%	0.022%	0.015%	0.010%	0.006%	0.004%	0.002%
-45.0%	21.0%	1.0%	1.297%	0.008	0.008%	0.641%	0.053%	0.045%	0.035%	0.026%	0.018%	0.012%	0.007%	0.004%	0.003%	0.001%
-48.4%	20.0%	0.9%	1.269%	0.008	0.006%	0.626%	0.050%	0.040%	0.030%	0.021%	0.013%	0.008%	0.005%	0.003%	0.001%	0.001%
-41.7%	20.0%	0.9%	1.269%	0.009	0.009%	0.630%	0.055%	0.048%	0.038%	0.029%	0.020%	0.013%	0.008%	0.005%	0.003%	0.002%
-41.7%	23.0%	0.9%	1.242%	0.008	0.010%	0.614%	0.052%	0.045%	0.037%	0.028%	0.021%	0.015%	0.010%	0.007%	0.004%	0.003%
-45.0%	24.0%	0.9%	1.228%	0.008	0.008%	0.603%	0.049%	0.041%	0.033%	0.025%	0.018%	0.013%	0.009%	0.006%	0.004%	0.002%
-45.0%	25.0%	0.9%	1.201%	0.008	0.008%	0.589%	0.047%	0.040%	0.032%	0.025%	0.018%	0.013%	0.009%	0.006%	0.004%	0.003%
-48.4%	23.0%	0.9%	1.187%	0.008	0.006%	0.582%	0.045%	0.037%	0.028%	0.021%	0.014%	0.010%	0.006%	0.004%	0.002%	0.001%
-41.7%	22.0%	0.9%	1.174%	0.008	0.009%	0.581%	0.050%	0.043%	0.035%	0.027%	0.020%	0.014%	0.009%	0.006%	0.004%	0.002%
-48.4%	17.0%	0.9%	1.160%	0.007	0.004%	0.575%	0.048%	0.038%	0.027%	0.017%	0.010%	0.005%	0.003%	0.001%	0.001%	0.000%

Fitted Distributions - LnNormal
Higher Skew

Exhibit 7

	E(X)	Mu	Sigma	Skew
Fitted	64.1%	(0.4500)	0.11	1.5

Fitted Original Distribution	**Ground-Up**	**Stop Loss Contract**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**64.1%**	**0.235%**	49.975%	4.785%	4.105%	2.858%	1.532%	0.629%	0.201%	0.052%	0.011%	0.002%	0.000%
Expected Loss On Line		**9.389%**	99.950%	95.707%	82.091%	57.152%	30.637%	12.572%	4.030%	1.040%	0.223%	0.041%	0.007%

Parameter Set Distributions	**Ground-Up**	**Stop Loss**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**64.6%**	**0.356%**	49.846%	4.610%	3.940%	2.847%	1.695%	0.868%	0.412%	0.193%	0.094%	0.048%	0.026%
Expected Loss On Line		**14.240%**	99.691%	92.198%	78.798%	56.937%	33.906%	17.353%	8.233%	3.868%	1.874%	0.955%	0.514%

Difference Parameter Set to Fitted	**0.73%**	**51.67%**	-0.26%	-3.67%	-4.01%	-0.38%	10.67%	38.03%	104.29%	272.02%	742.00%	2246.96%	7799.28%

Mu	Sigma	Freq	Relative Prob											
(0.4500)	0.0990	0.990%	2.787%	1.786%	0.005%	1.393%	0.135%	0.118%	0.081%	0.040%	0.014%	0.004%	0.001%	0.000%	0.000%	0.000%
(0.4500)	0.1045	0.960%	2.703%	1.733%	0.006%	1.351%	0.130%	0.113%	0.078%	0.040%	0.015%	0.004%	0.001%	0.000%	0.000%	0.000%
(0.4500)	0.1155	0.960%	2.703%	1.735%	0.007%	1.350%	0.128%	0.109%	0.077%	0.043%	0.019%	0.006%	0.002%	0.000%	0.000%	0.000%
(0.4500)	0.1100	0.950%	2.675%	1.716%	0.006%	1.337%	0.128%	0.110%	0.076%	0.041%	0.017%	0.005%	0.001%	0.000%	0.000%	0.000%
(0.4500)	0.1210	0.870%	2.449%	1.573%	0.007%	1.223%	0.115%	0.098%	0.069%	0.039%	0.018%	0.007%	0.002%	0.001%	0.000%	0.000%
(0.4500)	0.1265	0.850%	2.393%	1.538%	0.008%	1.195%	0.112%	0.094%	0.067%	0.039%	0.019%	0.008%	0.003%	0.001%	0.000%	0.000%
(0.4500)	0.0935	0.780%	2.196%	1.406%	0.003%	1.098%	0.107%	0.094%	0.064%	0.030%	0.010%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.4500)	0.0880	0.730%	2.055%	1.316%	0.003%	1.028%	0.101%	0.090%	0.060%	0.027%	0.008%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.4172)	0.1155	0.710%	1.999%	1.326%	0.008%	0.999%	0.097%	0.088%	0.067%	0.042%	0.021%	0.008%	0.003%	0.001%	0.000%	0.000%
(0.4172)	0.1100	0.670%	1.886%	1.250%	0.007%	0.943%	0.092%	0.084%	0.064%	0.039%	0.019%	0.007%	0.002%	0.001%	0.000%	0.000%
(0.4172)	0.1265	0.620%	1.745%	1.159%	0.008%	0.872%	0.084%	0.075%	0.058%	0.037%	0.020%	0.009%	0.003%	0.001%	0.000%	0.000%
(0.4500)	0.1320	0.610%	1.717%	1.105%	0.006%	0.857%	0.079%	0.067%	0.048%	0.029%	0.014%	0.006%	0.002%	0.001%	0.000%	0.000%
(0.4500)	0.1375	0.610%	1.717%	1.105%	0.006%	0.857%	0.079%	0.066%	0.048%	0.029%	0.015%	0.007%	0.003%	0.001%	0.000%	0.000%
(0.4839)	0.0935	0.600%	1.689%	1.046%	0.001%	0.844%	0.080%	0.065%	0.037%	0.015%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.4172)	0.1210	0.600%	1.689%	1.121%	0.007%	0.844%	0.082%	0.073%	0.056%	0.036%	0.018%	0.008%	0.003%	0.001%	0.000%	0.000%
(0.4839)	0.0990	0.570%	1.605%	0.994%	0.001%	0.802%	0.076%	0.060%	0.036%	0.015%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.4839)	0.1100	0.560%	1.577%	0.978%	0.002%	0.787%	0.073%	0.058%	0.036%	0.017%	0.006%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.4839)	0.1210	0.560%	1.577%	0.979%	0.003%	0.787%	0.071%	0.056%	0.036%	0.018%	0.007%	0.002%	0.001%	0.000%	0.000%	0.000%
(0.4839)	0.1045	0.550%	1.548%	0.960%	0.002%	0.774%	0.072%	0.057%	0.035%	0.015%	0.005%	0.001%	0.000%	0.000%	0.000%	0.000%

Fitted Distributions - LnNormal Exhibit 8

Five Years of Data

	E(X)	Mu	Sigma	Skew
Fitted	64.1%	(0.4500)	0.11	0.5

Fitted Original Distribution	**Ground-Up**	**Stop Loss Contract**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**64.1%**	**0.235%**	49.975%	4.785%	4.105%	2.858%	1.532%	0.629%	0.201%	0.052%	0.011%	0.002%	0.000%
Expected Loss On Line		**9.389%**	99.950%	95.707%	82.091%	57.152%	30.637%	12.572%	4.030%	1.040%	0.223%	0.041%	0.007%

Parameter Set Distributions	**Ground-Up**	**Stop Loss**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**65.7%**	**0.489%**	49.475%	4.393%	3.773%	2.837%	1.853%	1.121%	0.677%	0.424%	0.279%	0.192%	0.137%
Expected Loss On Line		**19.545%**	98.950%	87.862%	75.456%	56.743%	37.067%	22.414%	13.534%	8.488%	5.583%	3.840%	2.745%
Difference Parameter Set to Fitted	**2.43%**	**108.17%**	-1.00%	-8.20%	-8.08%	-0.72%	20.99%	78.29%	235.82%	716.34%	2407.96%	9339.89%	42089.69%

Mu	Sigma	Freq	Relative Prob											
-45.0%	9.9%	1.190%	1.246%	0.008	0.002%	0.623%	0.061%	0.053%	0.036%	0.018%	0.006%	0.002%	0.000%	0.000%	0.000%	0.000%
-45.0%	10.5%	1.140%	1.194%	0.008	0.002%	0.597%	0.058%	0.050%	0.034%	0.018%	0.007%	0.002%	0.000%	0.000%	0.000%	0.000%
-45.0%	8.3%	1.120%	1.173%	0.008	0.001%	0.586%	0.058%	0.052%	0.035%	0.015%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
-45.0%	9.4%	1.120%	1.173%	0.008	0.002%	0.586%	0.057%	0.050%	0.034%	0.016%	0.005%	0.001%	0.000%	0.000%	0.000%	0.000%
-45.0%	11.0%	1.100%	1.152%	0.007	0.003%	0.576%	0.055%	0.047%	0.033%	0.018%	0.007%	0.002%	0.001%	0.000%	0.000%	0.000%
-41.7%	9.4%	1.080%	1.131%	0.007	0.003%	0.566%	0.056%	0.052%	0.040%	0.023%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%
-48.4%	11.0%	1.030%	1.079%	0.007	0.001%	0.539%	0.050%	0.039%	0.024%	0.011%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
-45.0%	8.8%	1.000%	1.047%	0.007	0.001%	0.524%	0.051%	0.046%	0.031%	0.014%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
-48.4%	10.5%	0.960%	1.005%	0.006	0.001%	0.502%	0.047%	0.037%	0.023%	0.010%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
-41.7%	11.6%	0.940%	0.984%	0.007	0.004%	0.492%	0.048%	0.043%	0.033%	0.021%	0.010%	0.004%	0.001%	0.000%	0.000%	0.000%
-41.7%	11.0%	0.920%	0.964%	0.006	0.004%	0.482%	0.047%	0.043%	0.033%	0.020%	0.009%	0.004%	0.001%	0.000%	0.000%	0.000%
-41.7%	10.5%	0.900%	0.943%	0.006	0.003%	0.471%	0.046%	0.042%	0.033%	0.019%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%
-45.0%	7.7%	0.880%	0.922%	0.006	0.001%	0.461%	0.046%	0.041%	0.028%	0.011%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%
-45.0%	11.6%	0.880%	0.922%	0.006	0.002%	0.461%	0.044%	0.037%	0.026%	0.014%	0.006%	0.002%	0.001%	0.000%	0.000%	0.000%
-45.0%	13.2%	0.870%	0.911%	0.006	0.003%	0.455%	0.042%	0.036%	0.026%	0.015%	0.008%	0.003%	0.001%	0.000%	0.000%	0.000%
-45.0%	12.7%	0.830%	0.869%	0.006	0.003%	0.434%	0.041%	0.034%	0.024%	0.014%	0.007%	0.003%	0.001%	0.000%	0.000%	0.000%
-41.7%	8.8%	0.800%	0.838%	0.006	0.002%	0.419%	0.042%	0.039%	0.030%	0.017%	0.006%	0.001%	0.000%	0.000%	0.000%	0.000%
-48.4%	9.4%	0.780%	0.817%	0.005	0.001%	0.408%	0.039%	0.031%	0.018%	0.007%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%
-48.4%	9.9%	0.780%	0.817%	0.005	0.001%	0.408%	0.038%	0.031%	0.018%	0.008%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%

Higher Mean
Loss Ratio Distribution

Exhibit 9

	Actual LR	Actual Ln(LR)
1	60.0%	(0.5107)
2	70.9%	(0.3445)
3	74.4%	(0.2955)
4	64.9%	(0.4321)
5	60.0%	(0.5113)
6	70.9%	(0.3445)
7	85.4%	(0.1580)
8	77.4%	(0.2568)
9	66.9%	(0.4020)
10	70.9%	(0.3445)
Average	70.2%	(0.3600)
Stdev	0.0780	0.1100
Skew	0.5000	
Expected		70.2%

Min	72.5%
Max	75.0%
	Loss
	0.0%
	0.0%
	1.9%
	0.0%
	0.0%
	0.0%
	2.5%
	2.5%
	0.0%
	0.0%
Average	0.692%
Stdev	0.0112
Expected Loss On Line	27.7%
Fitted Expected Loss	0.769%
Fitted Expected Loss On Line	30.775%

Higher Standard Deviation
Loss Ratio Distribution

Exhibit 10

	Actual LR	Actual Ln(LR)
1	51.2%	(0.6695)
2	57.7%	(0.5504)
3	74.7%	(0.2920)
4	53.3%	(0.6299)
5	74.2%	(0.2985)
6	80.8%	(0.2137)
7	62.4%	(0.4724)
8	87.1%	(0.1380)
9	50.0%	(0.6933)
10	58.1%	(0.5424)
Average	64.9%	(0.4500)
Stdev	0.1324	0.2000
Skew	0.5000	
Expected		65.1%

Min	72.5%
Max	75.0%
	Loss
	0.0%
	0.0%
	2.2%
	0.0%
	1.7%
	2.5%
	0.0%
	2.5%
	0.0%
	0.0%
Average	0.887%
Stdev	0.0117
Expected Loss On Line	35.5%
Fitted Expected Loss	0.584%
Fitted Expected Loss On Line	23.379%

Higher Skewness
Loss Ratio Distribution

Exhibit 11

	Actual LR	Actual Ln(LR)
1	61.6%	(0.4842)
2	58.7%	(0.5327)
3	64.5%	(0.4379)
4	56.3%	(0.5746)
5	67.4%	(0.3946)
6	61.1%	(0.4921)
7	81.3%	(0.2071)
8	70.9%	(0.3437)
9	58.2%	(0.5411)
10	61.1%	(0.4921)
Average	64.1%	(0.4500)
Stdev	0.0747	0.1100
Skew	1.5000	
Expected		64.1%

Min	72.5%
Max	75.0%
	Loss
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	2.5%
	0.0%
	0.0%
	0.0%
Average	0.250%
Stdev	0.0079
Expected Loss On Line	10.0%
Fitted Expected Loss	0.235%
Fitted Expected Loss On Line	9.389%

Exhibit 12

Five Years of Data
Loss Ratio Distribution

	Actual LR	Actual Ln(LR)
1	58.1%	(0.5425)
2	67.3%	(0.3961)
3	74.2%	(0.2990)
4	56.6%	(0.5687)
5	64.2%	(0.4436)
6		
7		
8		
9		
10		
Average	64.1%	(0.4500)
Stdev	0.0712	0.1100
Skew	0.5000	
Expected		64.1%

Min	72.5%
Max	75.0%
	Loss
	0.0%
	0.0%
	1.7%
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
	0.0%
Average	0.165%
Stdev	0.0052
Expected Loss On Line	6.6%
Fitted Expected Loss	0.235%
Fitted Expected Loss On Line	9.389%

Fitted Distributions - LnNormal
Half Steps

Exhibit 13

	E(X)	Mu	Sigma	Skew
Fitted	64.1%	(0.4500)	0.11	0.5

		Stop Loss										
Fitted Original Distribution	**Ground-Up**	**Contract**										
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer	**64.1%**	**0.235%**	49.975%	4.785%	4.105%	2.858%	1.532%	0.629%	0.201%	0.052%	0.011%	0.002%	0.000%
Expected Loss on Line		**9.389%**	99.950%	95.707%	82.091%	57.152%	30.637%	12.572%	4.030%	1.040%	0.223%	0.041%	0.007%
Parameter Set Distributions	***Ground-Up***												
Loss Ratio Minimum Range		**72.5%**	0.0%	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%
Loss Ratio Maximum		**75.0%**	50.0%	55.0%	60.0%	65.0%	70.0%	75.0%	80.0%	85.0%	90.0%	95.0%	100.0%
Expected Loss to Layer Half Steps	**64.537%**	**0.345%**	49.855%	4.617%	3.946%	2.842%	1.676%	0.845%	0.393%	0.181%	0.086%	0.043%	0.022%
Exp Loss Original Steps	**64.536%**	**0.342%**	49.860%	4.622%	3.955%	2.850%	1.677%	0.841%	0.388%	0.176%	0.082%	0.039%	0.020%
Expected Loss On Line Half Steps		**13.790%**	99.709%	92.338%	78.922%	56.838%	33.522%	16.893%	7.867%	3.624%	1.719%	0.854%	0.447%
Expected Loss On Line Original Steps		**13.693%**	99.719%	92.445%	79.103%	56.994%	33.544%	16.812%	7.754%	3.516%	1.630%	0.788%	0.401%
Comparison with Larger Steps													
Diff in Exp Loss with Original Steps	**0.001%**	**0.002%**	-0.005%	-0.005%	-0.009%	-0.008%	-0.001%	0.004%	0.006%	0.005%	0.004%	0.003%	0.002%
Diff in ELOL's with Original Steps		**0.096%**	-0.010%	-0.106%	-0.182%	-0.156%	-0.022%	0.081%	0.113%	0.108%	0.089%	0.066%	0.046%
% Diff With Original Steps	**0.002%**	**0.703%**	-0.010%	-0.115%	-0.230%	-0.274%	-0.065%	0.480%	1.456%	3.070%	5.437%	8.372%	11.479%

ln x	y	Freq	Relative Prob										
(0.45000)	0.10725	0.0123	0.787%	0.50%	0.002%	0.393%	0.038%	0.033%	0.023%	0.012%	0.005%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.11000	0.0123	0.787%	0.50%	0.002%	0.393%	0.038%	0.032%	0.022%	0.012%	0.005%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.10450	0.012	0.767%	0.49%	0.002%	0.384%	0.037%	0.032%	0.022%	0.011%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.45000)	*0.10175*	*0.0115*	*0.735%*	*0.47%*	*0.001%*	*0.368%*	*0.036%*	*0.031%*	*0.021%*	*0.011%*	*0.004%*	*0.001%*	*0.000%*	*0.000%*	*0.000%*	*0.000%*
(0.43347)	0.09900	0.011	0.703%	0.46%	0.002%	0.352%	0.034%	0.031%	0.023%	0.012%	0.005%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.46681)	0.12100	0.0109	0.697%	0.44%	0.002%	0.348%	0.032%	0.026%	0.018%	0.010%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.43347)	0.09350	0.0107	0.684%	0.45%	0.001%	0.342%	0.034%	0.031%	0.022%	0.012%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.11275	0.0105	0.671%	0.43%	0.002%	0.336%	0.032%	0.027%	0.019%	0.010%	0.004%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.46681)	0.11000	0.0104	0.665%	0.42%	0.001%	0.332%	0.031%	0.026%	0.017%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.46681)	0.10450	0.0103	0.659%	0.42%	0.001%	0.329%	0.031%	0.026%	0.017%	0.008%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.09075	0.0103	0.659%	0.42%	0.001%	0.329%	0.032%	0.028%	0.019%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.11550	0.0103	0.659%	0.42%	0.002%	0.329%	0.031%	0.027%	0.019%	0.010%	0.005%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.46681)	0.10175	0.0102	0.652%	0.41%	0.001%	0.326%	0.031%	0.026%	0.017%	0.008%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.46681)	0.11550	0.0101	0.646%	0.41%	0.001%	0.323%	0.030%	0.025%	0.017%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.09625	0.01	0.639%	0.41%	0.001%	0.320%	0.031%	0.027%	0.019%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.43347)	0.10725	0.01	0.639%	0.42%	0.002%	0.320%	0.031%	0.028%	0.020%	0.011%	0.005%	0.002%	0.000%	0.000%	0.000%	0.000%
(0.45000)	0.09350	0.0099	0.633%	0.41%	0.001%	0.316%	0.031%	0.027%	0.018%	0.009%	0.003%	0.001%	0.000%	0.000%	0.000%	0.000%
(0.43347)	*0.09075*	*0.0099*	*0.633%*	*0.41%*	*0.001%*	*0.317%*	*0.031%*	*0.028%*	*0.021%*	*0.011%*	*0.004%*	*0.001%*	*0.000%*	*0.000%*	*0.000%*	*0.000%*

Figure 1

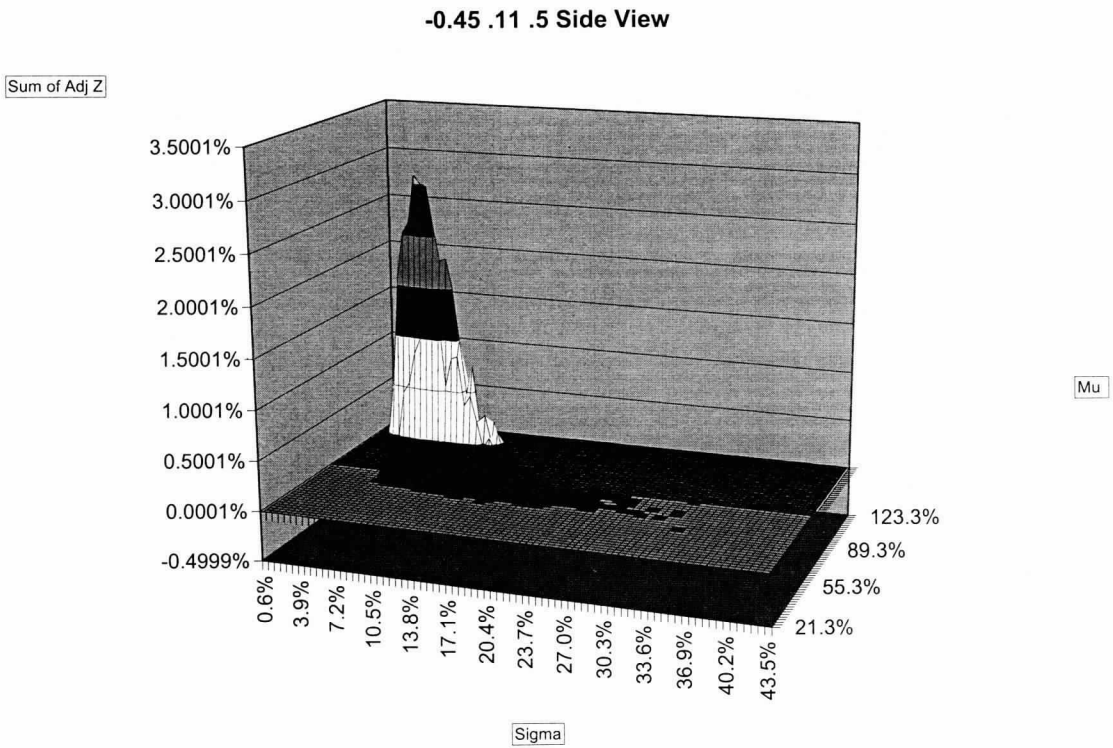

-0.45 .11 .5 Side View
Sum of Adj Z
3.5001%
3.0001%
2.5001%
2.0001%
1.5001%
1.0001%
0.5001%
0.0001%
-0.4999%
0.6%
3.9%
7.2%
10.5%
13.8%
17.1%
20.4%
23.7%
27.0%
30.3%
33.6%
36.9%
40.2%
43.5%
Sigma
123.3%
89.3%
55.3%
21.3%
Mu

Figure 2

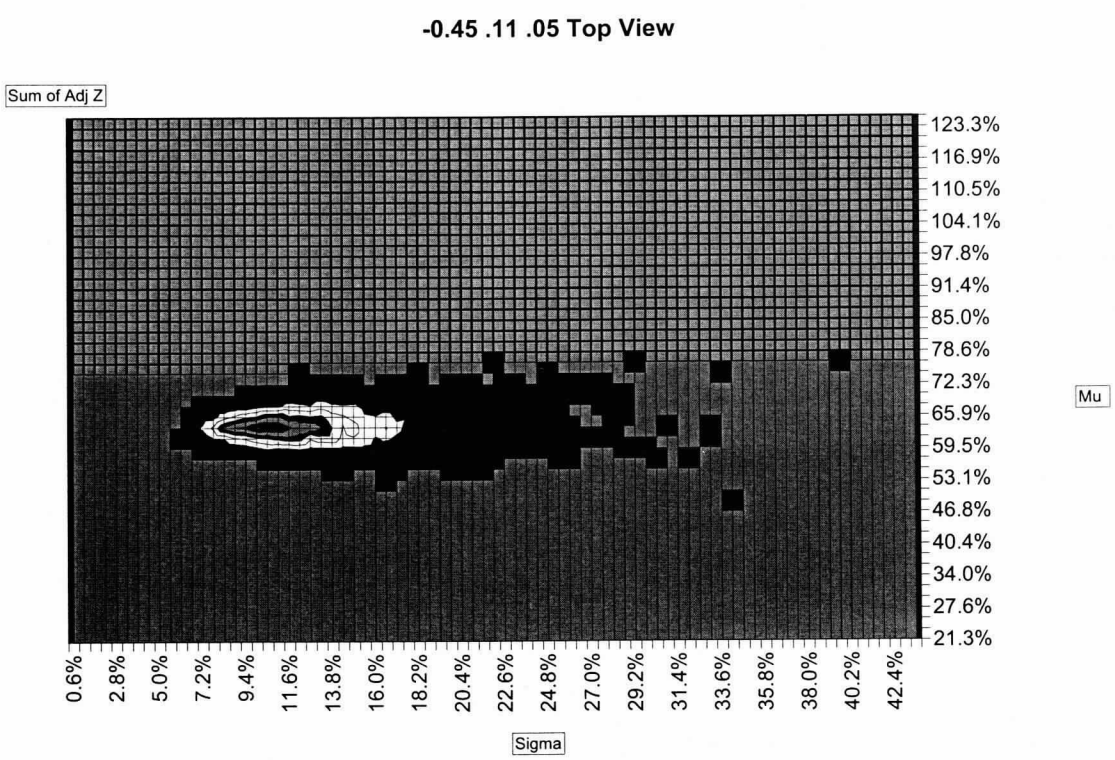

-0.45 .11 .05 Top View
Sum of Adj Z
123.3%
116.9%
110.5%
104.1%
97.8%
91.4%
85.0%
78.6%
72.3%
65.9%
59.5%
53.1%
46.8%
40.4%
34.0%
27.6%
21.3%
Mu
0.6%
2.8%
5.0%
7.2%
9.4%
11.6%
13.8%
16.0%
18.2%
20.4%
22.6%
24.8%
27.0%
29.2%
31.4%
33.6%
35.8%
38.0%
40.2%
42.4%
Sigma

Figure 3

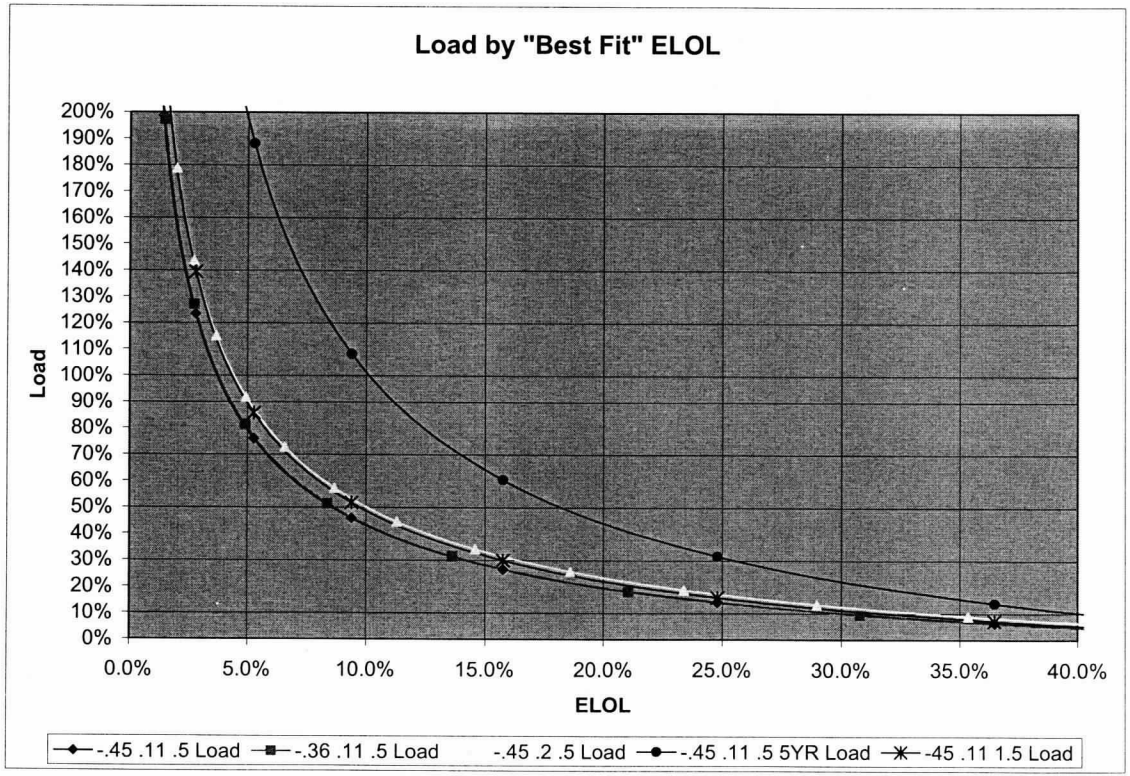

Load by "Best Fit" ELOL
Load
200%
190%
180%
170%
160%
150%
140%
130%
120%
110%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0.0%
5.0%
10.0%
15.0%
20.0%
25.0%
30.0%
35.0%
40.0%
ELOL
-.45 .11 .5 Load
-.36 .11 .5 Load
-.45 .2 .5 Load
-.45 .11 .5 5YR Load
-45 .11 1.5 Load

Figure 4

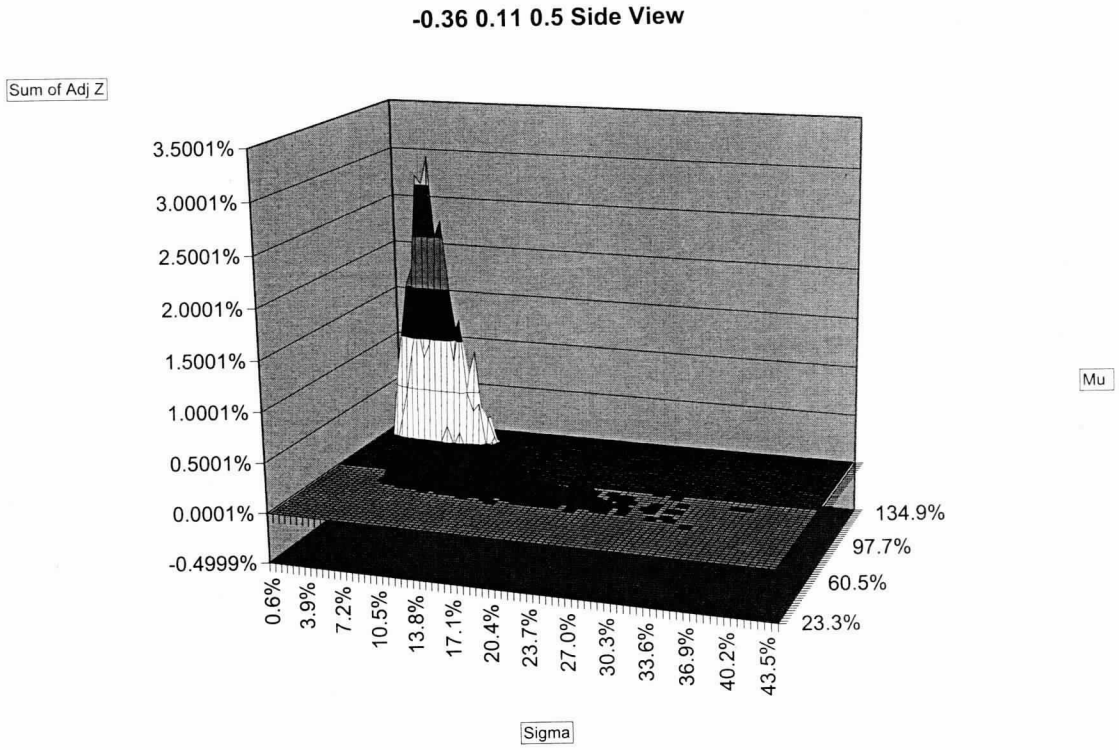

-0.36 0.11 0.5 Side View
Sum of Adj Z
3.5001%
3.0001%
2.5001%
2.0001%
1.5001%
1.0001%
0.5001%
0.0001%
-0.4999%
0.6%
3.9%
7.2%
10.5%
13.8%
17.1%
20.4%
23.7%
27.0%
30.3%
33.6%
36.9%
40.2%
43.5%
Sigma
134.9%
97.7%
60.5%
23.3%
Mu

Figure 5

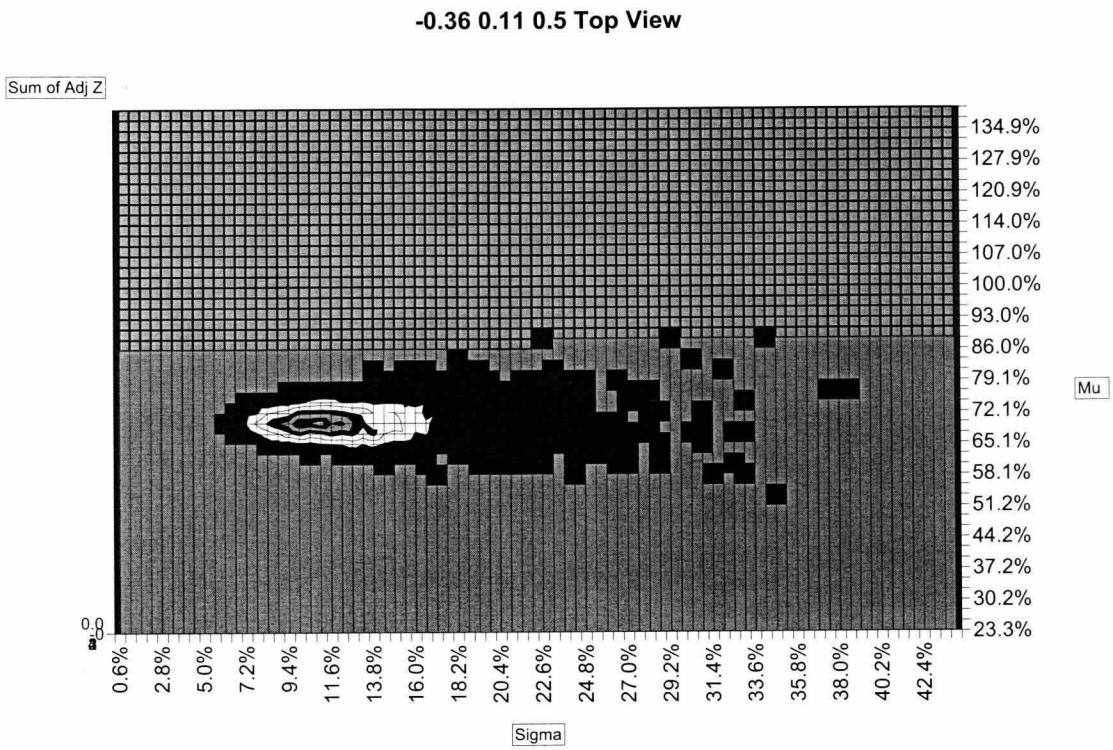

-0.36 0.11 0.5 Top View
Sum of Adj Z
134.9%
127.9%
120.9%
114.0%
107.0%
100.0%
93.0%
86.0%
79.1%
72.1%
65.1%
58.1%
51.2%
44.2%
37.2%
30.2%
23.3%
Mu
0.0
0.6%
2.8%
5.0%
7.2%
9.4%
11.6%
13.8%
16.0%
18.2%
20.4%
22.6%
24.8%
27.0%
29.2%
31.4%
33.6%
35.8%
38.0%
40.2%
42.4%
Sigma

Figure 6

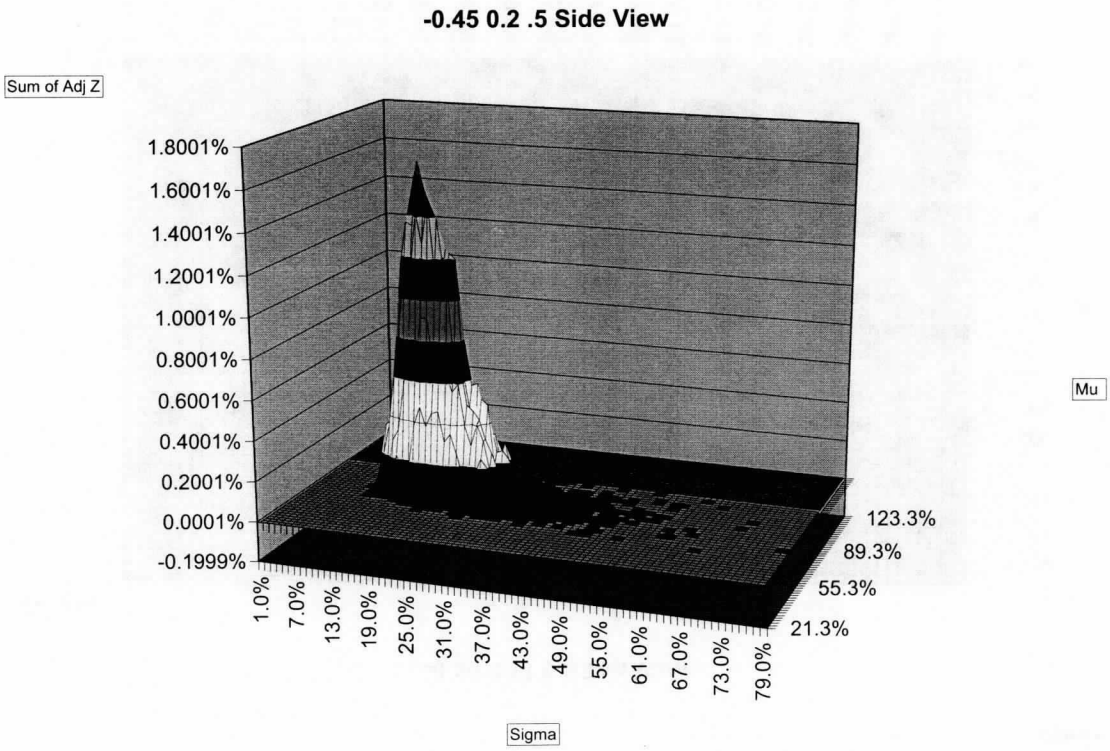

-0.45 0.2 .5 Side View
Sum of Adj Z
1.8001%
1.6001%
1.4001%
1.2001%
1.0001%
0.8001%
0.6001%
0.4001%
0.2001%
0.0001%
-0.1999%
1.0%
7.0%
13.0%
19.0%
25.0%
31.0%
37.0%
43.0%
49.0%
55.0%
61.0%
67.0%
73.0%
79.0%
Sigma
123.3%
89.3%
55.3%
21.3%
Mu

Figure 7

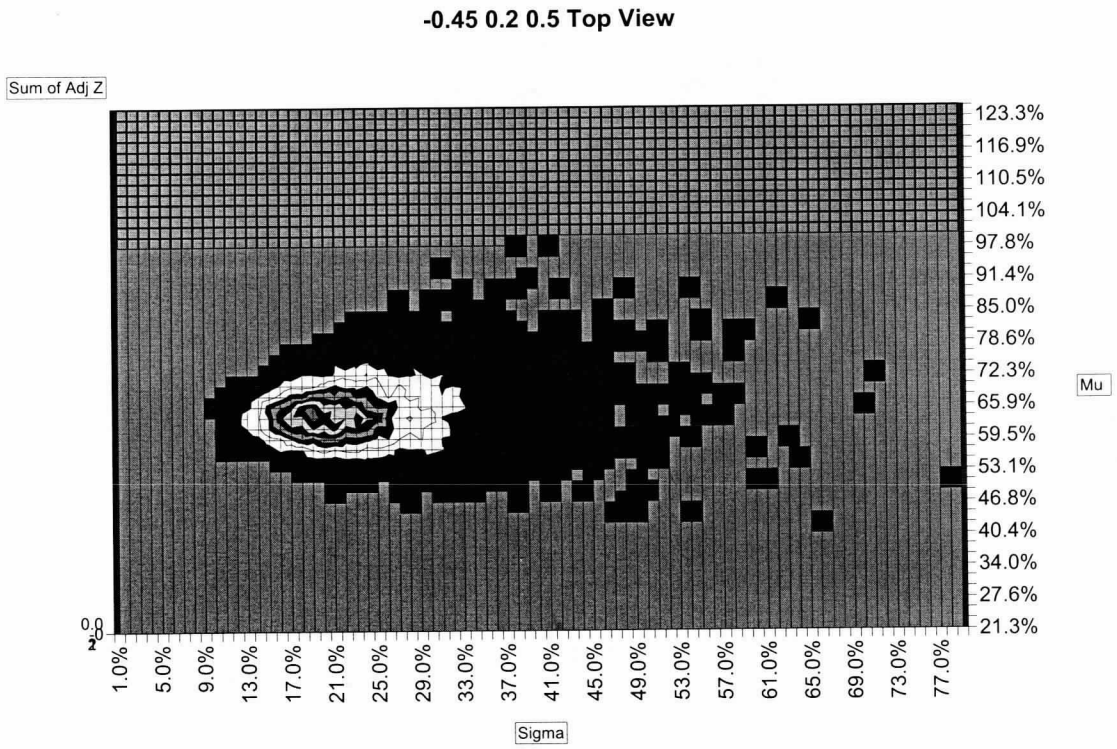
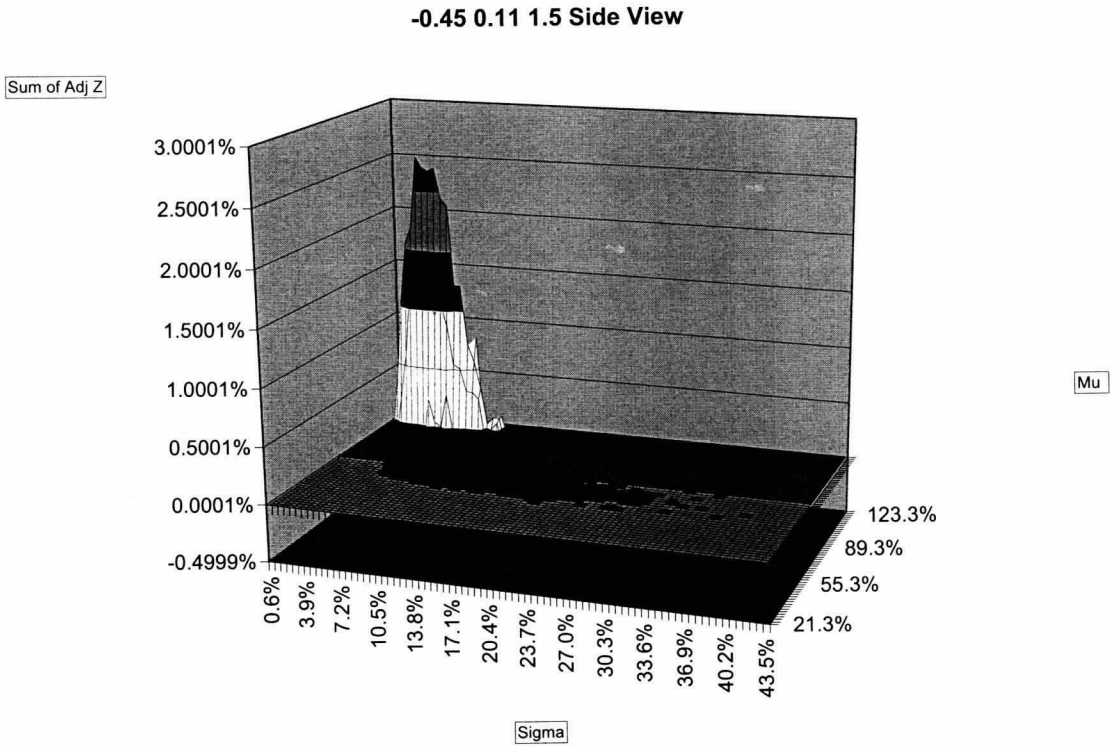

-0.45 0.2 0.5 Top View
Sum of Adj Z
123.3%
116.9%
110.5%
104.1%
97.8%
91.4%
85.0%
78.6%
72.3%
65.9%
59.5%
53.1%
46.8%
40.4%
34.0%
27.6%
21.3%
Mu
0.0
1.0%
5.0%
9.0%
13.0%
17.0%
21.0%
25.0%
29.0%
33.0%
37.0%
41.0%
45.0%
49.0%
53.0%
57.0%
61.0%
65.0%
69.0%
73.0%
77.0%
Sigma

Figure 8


-0.45 0.11 1.5 Side View
Sum of Adj Z
3.0001%
2.5001%
2.0001%
1.5001%
1.0001%
0.5001%
0.0001%
-0.4999%
0.6%
3.9%
7.2%
10.5%
13.8%
17.1%
20.4%
23.7%
27.0%
30.3%
33.6%
36.9%
40.2%
43.5%
Sigma
123.3%
89.3%
55.3%
21.3%
Mu

210

Figure 9

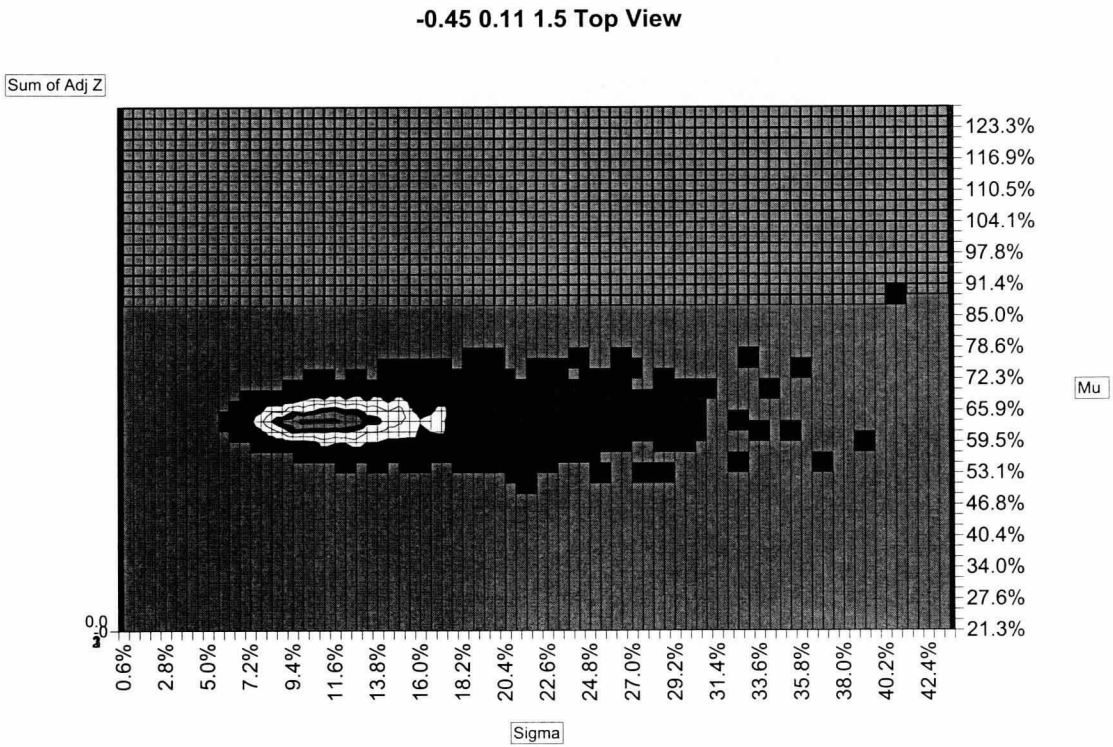

-0.45 0.11 1.5 Top View
Sum of Adj Z
123.3%
116.9%
110.5%
104.1%
97.8%
91.4%
85.0%
78.6%
72.3%
65.9%
59.5%
53.1%
46.8%
40.4%
34.0%
27.6%
21.3%
Mu
0.0
0.6%
2.8%
5.0%
7.2%
9.4%
11.6%
13.8%
16.0%
18.2%
20.4%
22.6%
24.8%
27.0%
29.2%
31.4%
33.6%
35.8%
38.0%
40.2%
42.4%
Sigma

Figure 10

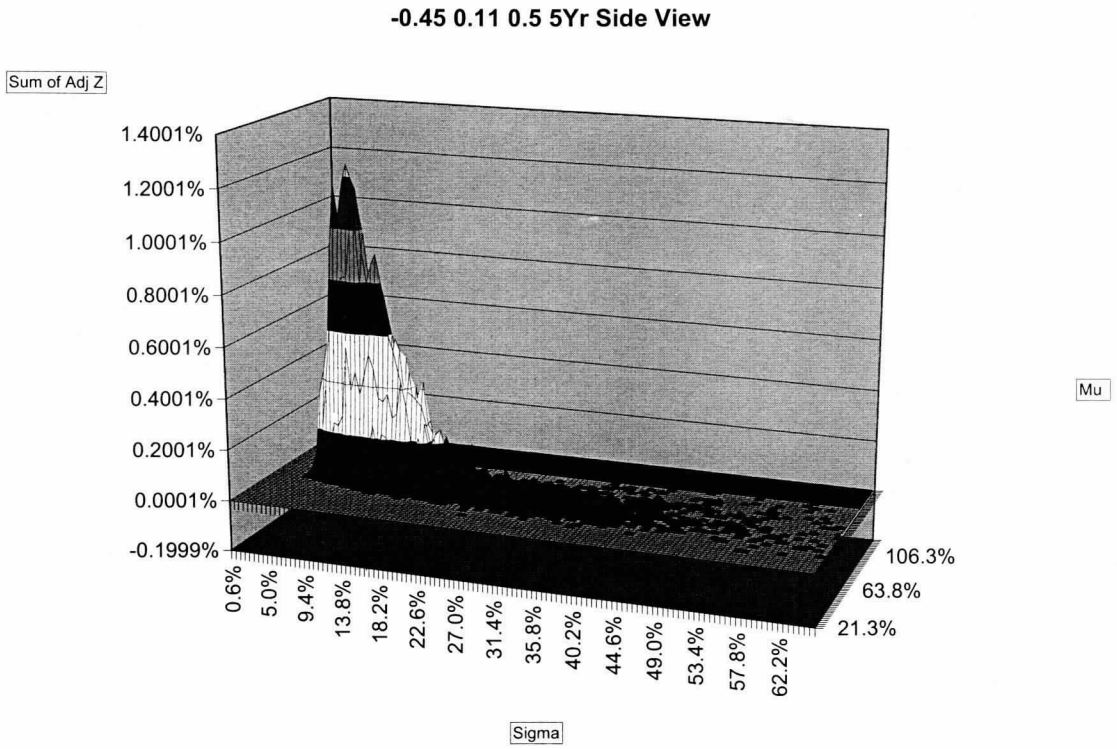
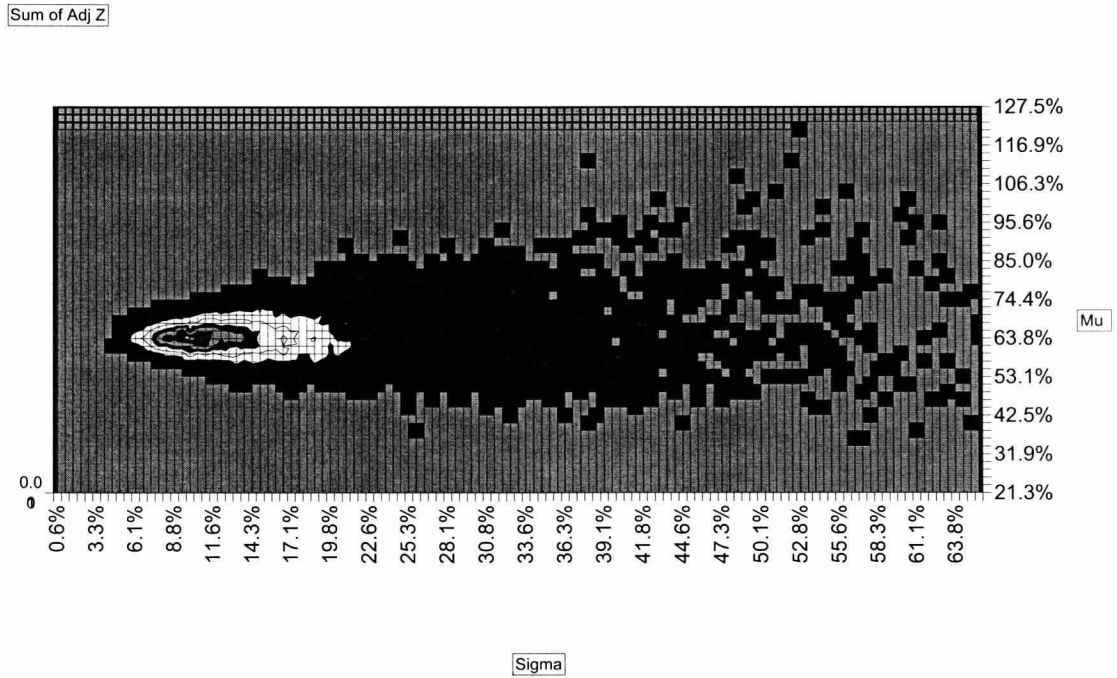

-0.45 0.11 0.5 5Yr Side View
Sum of Adj Z
1.4001%
1.2001%
1.0001%
0.8001%
0.6001%
0.4001%
0.2001%
0.0001%
-0.1999%
0.6%
5.0%
9.4%
13.8%
18.2%
22.6%
27.0%
31.4%
35.8%
40.2%
44.6%
49.0%
53.4%
57.8%
62.2%
Sigma
106.3%
63.8%
21.3%
Mu

Figure 11

-0.45 0.11 0.5 5Yr Top View



Quantifying Correlated Reinsurance Exposures with Copulas

Gary G. Venter, FCAS, MAAA

Quantifying Correlated Reinsurance Exposures with Copulas

Gary G Venter, Guy Carpenter Instrat

Copulas provide a convenient way to represent joint distributions. In fact the joint distribution function can be expressed as the copula function applied to the separate individual distributions. That is, $F(x_1, x_2, \ldots x_m) = C[F_1(x_1), F_2(x_2), \ldots F_m(x_m)]$ where C is the copula function. Background information on copulas is covered in my Proceedings paper *Tails of Copulas*, and will be largely assumed here.

That paper gave illustrations of bivariate copulas, most of which do not extend well into higher dimensions. For a multivariate copula for reinsurance related variates you would like to be able to feed in a correlation matrix of the variates as well as have some control over the degree of correlation in the tails of the distributions. Often more than two related variates are needed, such as losses in different lines of insurance.

This paper focuses on the t-copula, which meets these minimum requirements, but just barely. You can input a correlation matrix and you do have control over the tail behavior, but you only have one parameter to control the tail, so all pairs of variates will have tail correlation that is determined by that parameter. The normal copula is a limiting case, in which the tails are ultimately uncorrelated if you go out far enough.

The structure of the paper is to jump right in to a discussion of the t-copula in the bivariate case, then extend this to higher dimensions. A trivariate example is given using cat model output for three lines of insurance. Methods for selecting parameters and testing goodness of fit are discussed in this context.

ACKNOWLEDGEMENT

Much credit must go to Andrei Salomatov for solving the calculation issues for certain extreme cases discussed below.

The Bivariate t-Copula

The bivariate t-copula has two parameters that control the tail dependence and the degree of correlation separately.

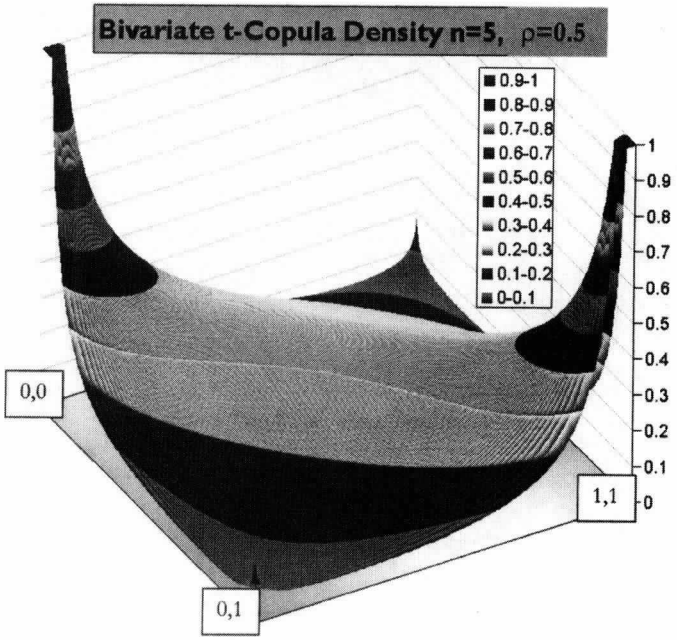
The t-distribution with n degrees of freedom is defined by:

$$f_n(x) = K_1(1+x^2/n)^{-(n+1)/2}, \text{ with } K_1=\Gamma(1/2+n/2)(n\pi)^{-1/2}/\Gamma(n/2).$$

Here n is often an integer, but doesn't have to be. The distribution is symmetric around zero and can be calculated by:

$$F_n(x) = \tfrac{1}{2} + \tfrac{1}{2}\,\text{sign}(x)\text{betadist}[x^2/(n+x^2), \tfrac{1}{2}, n/2],$$

where betadist defines the beta distribution. as in Excel.

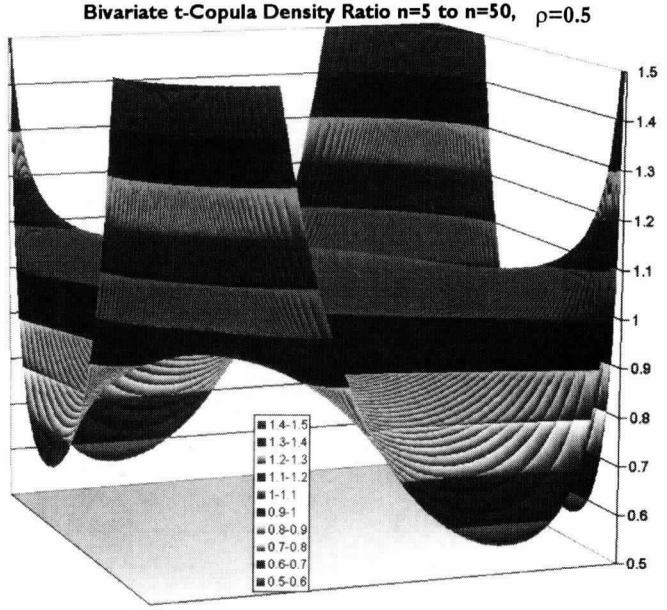


An example of the density of the t-copula is graphed above. The density of the bivariate t-copula with n degrees of freedom and the correlation parameter ρ is defined as:

$$c(u,v;\, n,\rho) = K_2[(1+s^2/n)(1+t^2/n)]^{(n+1)/2}\{1+[s^2-2\rho st+t^2]/[(1-\rho^2)n]\}^{-1-n/2}$$

with $K_2 = \frac{1}{2}[\Gamma(n/2)/\Gamma(0.5+n/2)]^2 n(1-\rho^2)^{-1/2}$ and $s=F_n^{-1}(u)$, $t=F_n^{-1}(v)$.

The inverse t-distribution needed for this can be calculated with an inverse beta by $s = \text{sign}(u-\frac{1}{2})n^{1/2}[-1+1/\text{betainv}(|2u-1|,\frac{1}{2}, n/2)]^{-1/2}$.



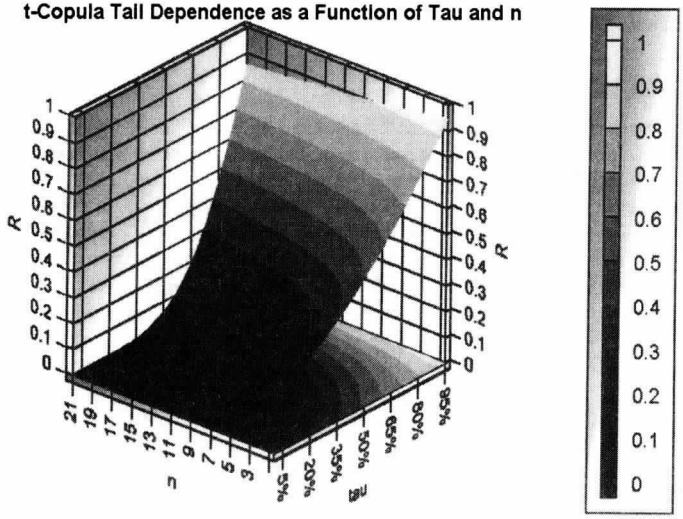
The concentrations of probability near [0,0] and [1,1] are seen in many copulas, but the smaller concentrations around [0,1] and [1,0] are more unusual. The above graph shows the ratio of densities for n=5 to n=50. The latter is similar to the normal copula density, which approaches zero

at [0,1] and [1,0]. Thus the density ratio is highest in those regions, even though it is well above unity around [0,0] and [1,1]. With a given correlation parameter or matrix, the linear correlation and Kendall's τ for the t-copula are the same for any n as for the normal copula ($n \to \infty$). The lower values of n produce greater upper and lower tail dependence with the same overall correlation essentially because they put more weight in all the corners. The additional weight in the off-diagonal corners cancels out the additional tail dependence, keeping the overall correlation the same.

Kendall's τ is related to ρ by $\tau = (2/\pi)\arcsin(\rho)$. Also the right tail dependence measure R, defined as $\text{limit}_{z \to 1} \Pr(U>z \mid V>z)$, is given by:

$$R/2 = 1 - F_{n+1}\{[(n+1)(1-\rho)/(1+\rho)]^{0.5}\}.$$

Thus R can be expressed as a function of τ and n, as graphed below. Even zero τ can give a positive tail dependence with this copula. The tail dependence can approach zero for any τ by taking n large, thus approximating the limit of the normal copula, which has tail dependence of zero.



The Multivariate t-Copula

To define the multivariate copula, suppose there are m variates, and **u** is a vector of m probability values (numbers in [0,1]). Let **s** be the vector of the univariate t-quantiles of **u** with n degrees of freedom, that is $s=F_n^{-1}(u)$ for each element of **s** and **u**. Also let Σ be an m x m correlation matrix with determinant d. Then the m-dimensional t-copula has density:

$$c(u; n, \Sigma) = K_m[\Pi_{i=1}^{m}(1+s_i^2/n)]^{(n+1)/2}(1+\mathbf{s}'\Sigma^{-1}\mathbf{s}/n)^{-(m+n)/2}$$

where $K_m = \Gamma[(m+n)/2][\Gamma(n/2)]^{m-1}[\Gamma(½ +n/2)]^{-m}d^{-1/2}$.

By starting with a Kendall's τ coefficient matrix T, the correlation matrix needed here can be specified by $\Sigma = \sin(T\pi/2)$. Thus this copula has complete flexibility in its correlation structure. However there is only one n used, so the tail dependence will be determined by that n for all pairs of variates. In the graph above, all the tail-dependence measures for all pairs of variates would fall on the same vertical cross section, determined by the value of n used for the copula. Thus the pairs with higher τ will have higher R as well.

The univariate or multivariate t distribution can be characterized (and simulated) by a (possibly multivariate) normal distribution divided by a multiple of the square root of an independent univariate chi-squared distribution. When generating it in this way, if a low draw comes up for the chi-squared variate, large values of the t variate can be produced, even if the normal values were not particularly large. In the multivariate case then all the t variates can be jointly large even if they were not originally correlated. That illustrates why the tail-dependence can be somewhat high even with zero τ. An example might be where the reciprocal of the t variate represents the inflation rate, which hits all the lines. This effect is sometimes called a common shock, i.e., the common shock of a large inflation rate can induce a correlation among otherwise independent lines.

More precisely, to generate a vector of probabilities from the multi-variate t-copula, first generate a multi-variate normal vector with the same correlation matrix, then divide it by $(y/n)^{0.5}$ where y is a number simulated from a chi-squared distribution with n degrees of freedom. This gives a t-

distributed vector, and the t-distribution F_n can then be applied to each element to get the probability vector.

The ratio y/n is a scale transform of the chi-squared variate, so is a gamma variate. If the gamma density is parameterized to be proportional to $x^{\alpha-1}e^{-x/\beta}$, then y/n has parameters $\beta = 2/n$ and $\alpha = n/2$. This is a distribution with mean 1. It can be simulated easily if an inverse gamma function is available, as in some spreadsheets.

Because a power of the gamma deviate is a divisor, a factor is being applied that is actually inverse transformed gamma distributed. The inverse transformed gamma distribution in α, τ, θ has density proportional to $\exp(-(\theta/x)^{\tau})/x^{\alpha+1}$. The factor $(n/y)^{0.5}$ applied to the normal variates is distributed inverse transformed gamma in $\alpha = n$, $\tau = 2$, and $\theta = (n/2)^{1/2}$. This has a mean greater than unity and an inverse power tail with power n, and so is a heavy-tailed distribution. Especially when n is small, this gives the possibility of large values of the factor occasionally being applied to all the normal draws, giving simultaneous large values of all the variates.

Example – Hurricane Losses

Parameter estimation issues and applications can be illustrated by a sample of losses simulated from a hurricane model. The simulation generated losses under three lines of insurance: residential property (R), commercial property (C) and automobile (A). Naturally these are highly correlated losses, as hurricane losses from a stronger storm tend to be higher in all three lines. The strength of the storm could be considered to be the common shock that correlates all the lines. Having a large generated sample like this does not require a fitted copula to be useful in loss estimation, so in practice there would be little need to fit a copula to it. It is a useful dataset for illustrating fitting concepts, however.

The empirical trivariate copula can be calculated at any 3-vector of probabilities by counting the proportion of the sample triplets of empirical probabilities that are less in each index. Each of the three bivariate empirical copulas from the three pairs of variables can be calculated similarly. The averages of the bivariate copulas give estimates of the τ correlations by the relationship $\tau = 4E(C) - 1$, where E(C) is the ex-

pected value of the copula.. This scaling of the mean value of the copula can be extended to define higher dimensional analogues of τ by requiring that $\tau = 0$ for the independence case and $\tau = 1$ for perfect correlation. The scaling for m-dimensions that does this is $\tau = [2^m E(C) - 1]/[2^{m-1} - 1]$. (There are other possible multi-variate extensions of τ, but they will not be used here.) For the hurricane data, these τ's are:

	AC	AR	RC	ARC
τ	82.4%	84.4%	87.6%	84.8%

The bivariate τ's provide estimates of the correlation ρ for each bivariate copula, and thus for the correlation matrix for the trivariate copula, using $\rho = \sin(\pi\tau/2)$. For the sample, these are:

	AC	AR	RC
ρ	.96	.97	.98

To estimate n, the tail behavior is key. One avenue might be to estimate R, the limit $_{z \to 1} \Pr(U>z \mid V>z)$. However this is difficult to estimate from data because the function R(z), defined as $\Pr(U>z \mid V>z)$, can drop rapidly for z near 1, and there is less and less data to use the closer z gets to 1.

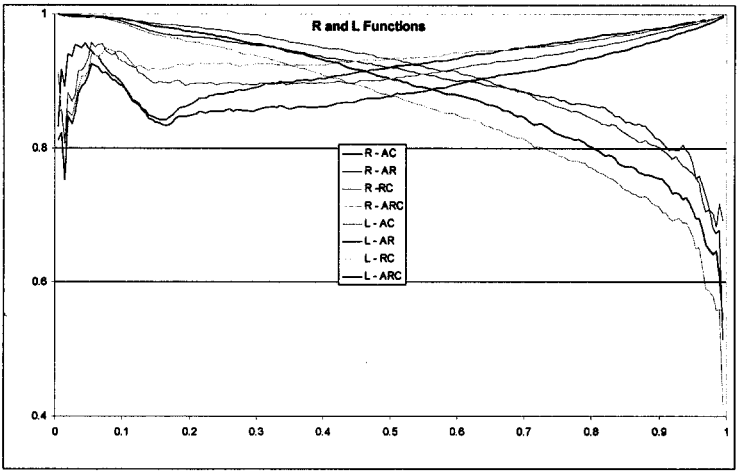
Note that $R(z) = \Pr(U>z \ \& \ V>z)/\Pr(V>z)$. Since $\Pr(V>z) = 1 - z = \Pr(U>z)$, U and V can be switched in the definition of R(z). A similar concept can be defined for the multivariate copula:

$$R(z) = \Pr(U>z \ \& \ V>z \ \& \ W>z)/z = \Pr(U>z \ \& \ V>z \mid W>z)$$

Because of the symmetry in the first equation, U, V, and W can be swapped around at will in the second equation. This function provides a measure of the overall tail dependency of the three variates, and it can be generalized to higher dimensions. A similar tail dependency function can be defined for the left tail:

$L(z) = \Pr(U<z \ \& \ V<z \ \& \ W<z)/z = C(z,z,z)/z$, and similarly in the bivariate case. The empirical versions of these functions are graphed below. From the graph, the right and left tail functions are clearly not

symmetrical. This would rule out the t-copula, which is. However most issues concern the large loss cases, so a copula approximating the right side would be most appropriate, and the t-copula might work for this.

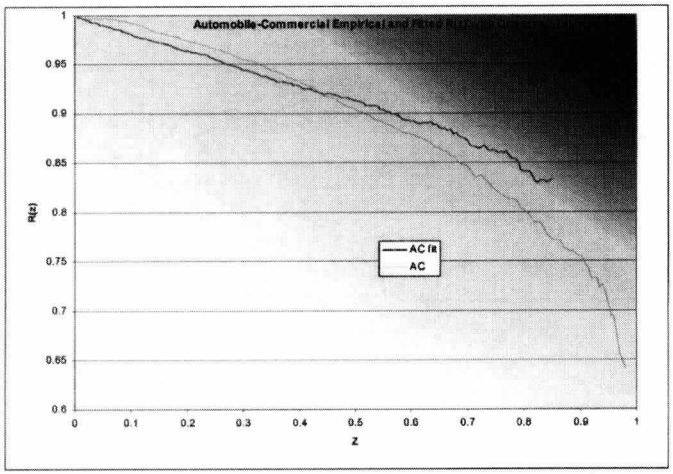


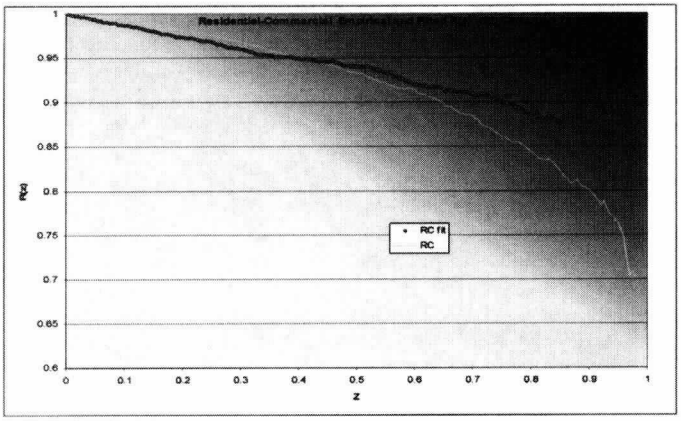
One possibility would be to estimate the t-copula correlations and degrees of freedom by maximum likelihood. The likelihood function for a parametric copula at a point in the sample is the density of the parametric copula computed at the empirical copula vector for that point, so can be readily calculated.

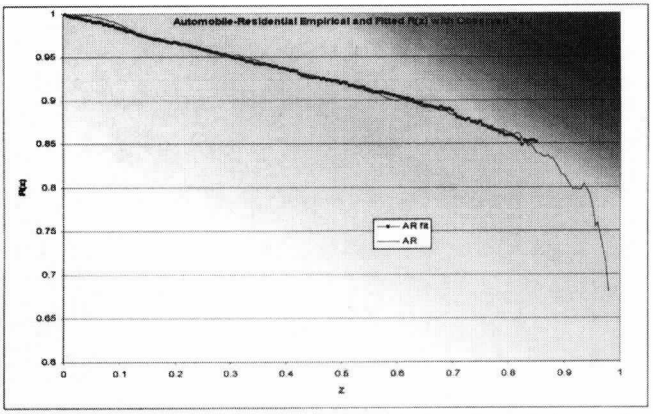
However in this case, MLE is not likely to give the intended fit, in that it would be affected by the smaller claims that do not appear to mirror the large claims. So the sample correlations come back as a starting reference point, even though they use the whole distribution.

To test how well the sample correlations match the larger losses, a simulation can be performed with the sample correlations and a selected n, and the simulated R(z) compared to the sample's. The arbitrary choice of degrees of freedom most affects the extreme percentiles, so this comparison was initially cut off at the 85th percentile, with a selected n=20. The bivariate comparisons are graphed below.

The AR R(z) comparison is very close by this measure, while the other two pairs do not fit very well.

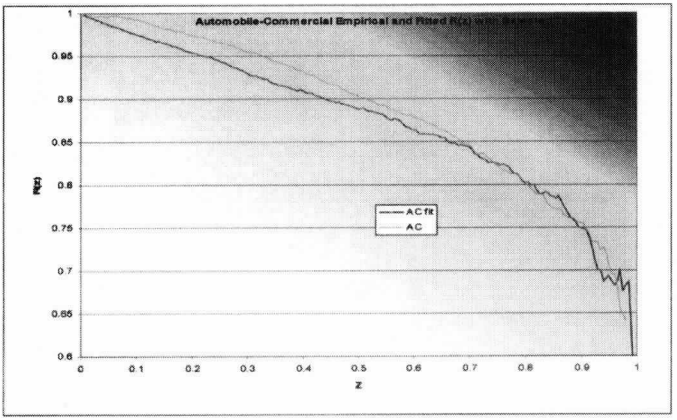


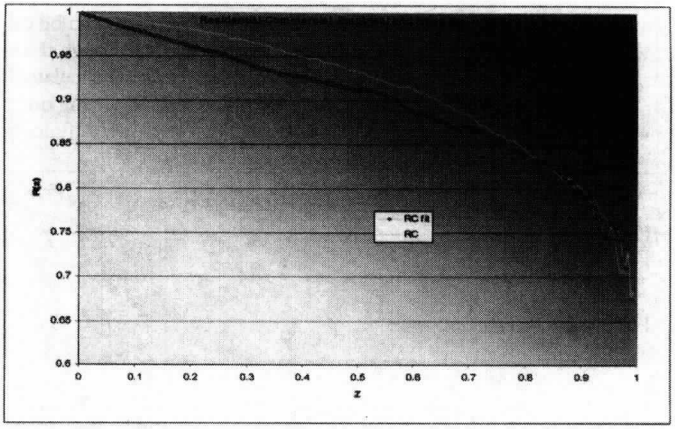


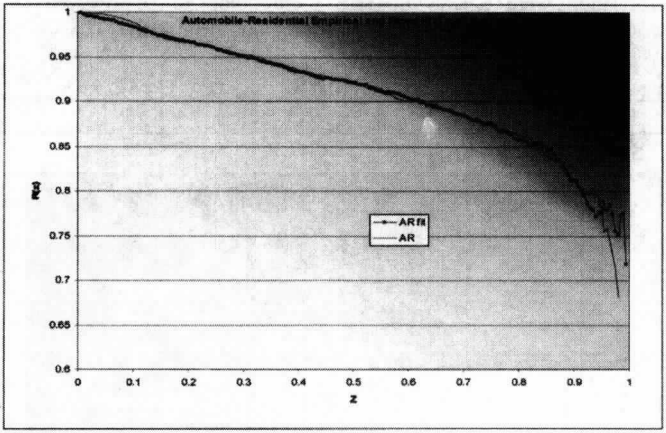


To see how much this is influenced by the choice of correlations, a few correlations were tested by this same methodology to see how well they work. The selections and fits are shown below, again with n=20.

	AC	AR	RC
Sample ρ	.96	.97	.98
Selected ρ	.94	.97	.96

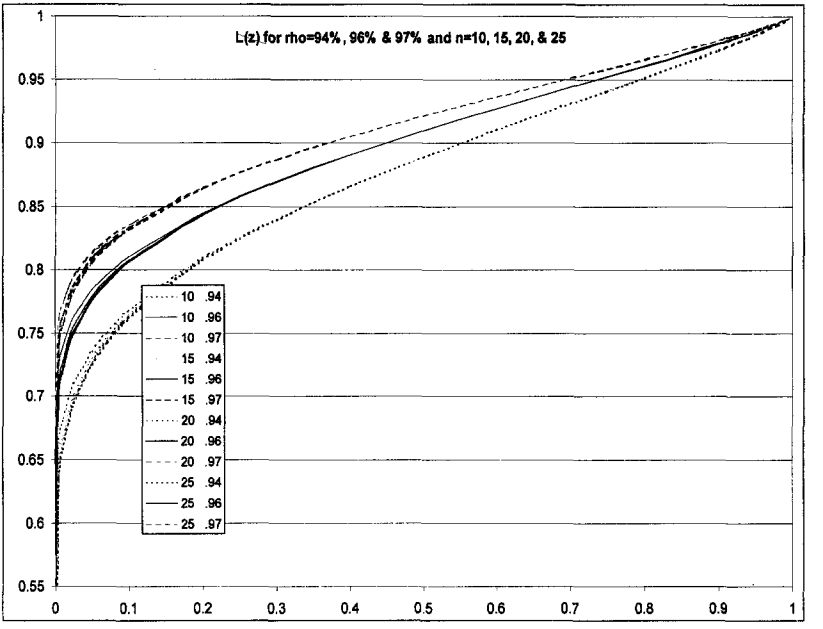






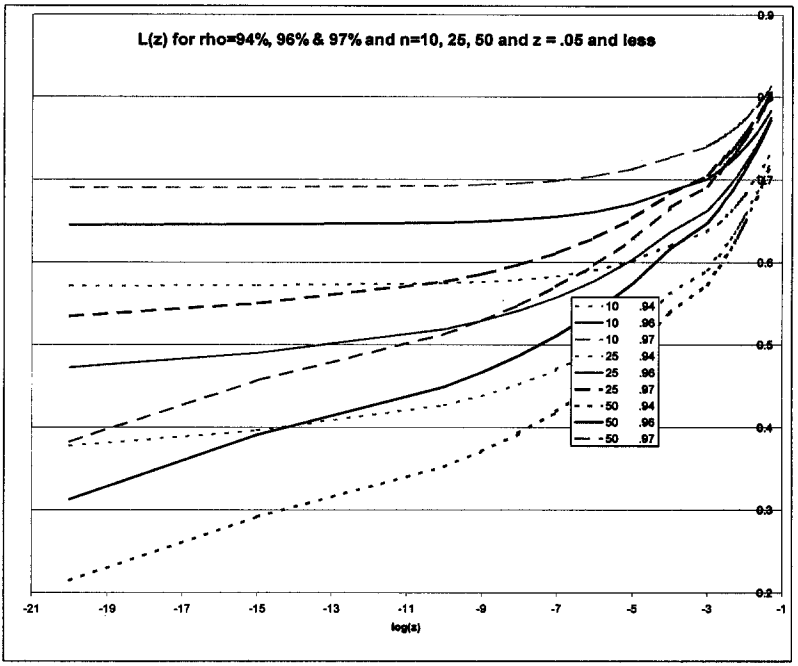
While there are still some issues with the fits for smaller values of z, they are much better for the large losses, as intended. This case was not cut off at 85%, and some simulation instability shows for the larger values of z.

Due to the symmetry of the t-copula, $R(z) = L(1 - z)$, and L can be calculated directly from the copula function. However, even though the t-copula is easy to simulate, the copula function is difficult to calculate, as the integration is difficult near 0. Numerical integration relying on simulation is often used for this. The graph below shows the L function for the ρ's fit above for a few values of n using this approach.



It is only for small values of z that n makes a difference in the L and R functions, at least for these large values of ρ. The ρ itself does affect the functions for all values of z.

The graph below looks at L(z) for $z \leq 0.05$ for these same ρ's and a few n's. It is on a log scale to illustrate the behavior of L for very small z's.



For z in this range, which is shown down to ln(z) = -20, n is at least as important as ρ in influencing the value of L(z). Also, the function declines very slowly even for n=50. With these values of z the tails become less determined by ρ, but only in the very extreme tail, beyond the area of practical concern.

To select a value of n, the empirical R functions were evaluated at z=0.005 and z=0.01, and n's sought to best match. For the RC pair with ρ = 96%, the best n was 41.5. For the other cases, n=500 worked as well as anything, suggesting a normal case. The target and fitted R(z)'s are shown for each pair in the table below. The RC fit is best. The other two fit ok at z = 0.01, but drop off for smaller z. That could

be a sample size problem at this level.

Target/Fit	94% (AC)	97% (AR)	96% (RC)
0.005	.54 / .61	.52 / .68	.693 / . 695
0.01	.61 / .64	.68 / .70	.718 / .715

Only one value of n is used in the t-copula, so a compromise value has to be selected. Perhaps a value near 42 would be appropriate. This works for RC but imposes too much tail association for the other pairs. This is only in the extreme tail, however, so might not be problematic.

Summary

The functional form of the t-copula is somewhat complicated, but most of the key functions are readily available in spreadsheets and statistical packages. Simulating samples is quite easy, as this just uses a simple adjustment to normal copula samples.

Estimating parameters from data is more problematic. If the data is symmetric, maximum likelihood would be a good choice. In that case, a comparison of the empirical and fitted R and L functions could be used to evaluate goodness of fit.

When the right and left tails are quite different the t-copula would not usually be indicated, but if only the right tail behavior is important in practice, a fit to that could be sought. Finding parameters that match the empirical and fitted R function is a reasonable way to do that. In the sample data reviewed, finding a match for the correlation matrix was relatively straightforward, but finding the best n was more difficult. For the high correlations found in this sample, different values of n affected R(z) only in the very extreme tail – even beyond where most reinsurance interest would be. Since that is where the data is most scarce, reliable fits are difficult. However the choice of n is not too critical for the same reason.

The main practical obstacle to the use of the t-copula is that there is only one parameter – n – to control tail association, and different pairs of variates might have different indicated n's. Computationally the biggest problem is calculating C for extreme values. This would be necessary only for trying to fit parameters to the extreme tail, however.

On the Optimality of Multiline Excess of Loss Covers

Jean-François Walhin

On the optimality of multiline excess of loss covers

J.F. Walhin

Institut des Sciences Actuarielles, Université Catholique de Louvain

Secura Belgian Re

Jean-François Walhin
Secura Belgian Re
Avenue des Nerviens, 9-31
B-1040 Bruxelles
Belgique
phone : + 32 2 504 82 22
fax : + 32 2 504 82 00
e-mail : jfw@secura-re.com

ON THE OPTIMALITY OF MULTILINE EXCESS OF LOSS COVERS

ABSTRACT

It is well known that diversifying the risk between independent policies reduces the total risk in the sense that less deviations around the aggregate mean loss are expected. In other words, less capital has to be allocated due to the diversification effect.
The same effect can be obtained when an insurance company buys an excess of loss cover. Instead of buying independently covers for different lines of business, it is intuitively acceptable to believe that the insurance company has interest in diversifying by buying a multiline excess of loss cover.
In the present paper I show how to deal with the dependencies induced by such a model and using some risk measures we show on a numerical example the optimality of the multiline agreement.

KEYWORDS

Multivariate Panjer's algorithm, multiline excess of loss cover, standard deviation, Wang Transform, optimal reinsurance.

BIOGRAPHY

Jean-François Walhin is R&D Manager at Secura Belgian Re which he joined in 1999 after having worked for three years for a primary insurer. He is also a visiting Professor at the Catholic University of Louvain-la-Neuve (UCL). Jean-François is a civil engineer and an actuary and holds a PhD in science from the UCL. He is the author of 20 papers that have appeared in actuarial journals. Jean-François is a Fellow of the ARAB-KVBA, the Belgian Association of Actuaries.

ACKNOWLEDGMENTS

I would like to thank Bryan Ware and Kenneth Ruppert for helpful remarks that led to a better presentation of the paper.

1. INTRODUCTION

Multiline excess of loss covers are introduced in Ribeaud (2000).
Walhin (2002) introduced the practical pricing of multiline excess of loss covers. To keep things simple, we will assume that we have two lines of business : Fire and MTPL. Let us define

- X_i^{Fire} as the i^{th} claim amount of type Fire,
- X_i^{MTPL} as the i^{th} claim amount of type Motor Third Part Liability (MTPL in short).

It is assumed that the X_i^{Fire}'s are independent and identically distributed as well as the X_i^{MTPL}'s. X_i^{Fire}'s and X_i^{MTPL}'s are assumed to be mutually independent. We also define

- N as the number of claims of type Fire,
- M as the number of claims of type MTPL.

We assume that N and M are independent and that N and the X_i^{Fire}'s on the one hand and M and the X_i^{MTPL}'s on the other hand are also independent.
Let us define the liability of the excess of loss reinsurer for each claim :

$$\begin{aligned} R_i^{Fire} &= \min(L^{Fire}, \max(0, X_i^{Fire} - D^{Fire})), \\ R_i^{MTPL} &= \min(L^{MTPL}, \max(0, X_i^{MTPL} - D^{MTPL})). \end{aligned}$$

where

1. D^{Fire} is the deductible for fire claims
2. L^{Fire} is the limit for fire claims
3. D^{MTPL} is the deductible for MTPL claims
4. L^{MTPL} is the limit for MTPL claims

Let us define the liability of the ceding company for each claim :

$$\begin{aligned} C_i^{Fire} &= X_i^{Fire} - R_i^{Fire}, \\ C_i^{MTPL} &= X_i^{MTPL} - R_i^{MTPL}. \end{aligned}$$

Let us define the aggregate liability of the reinsurer for each line :

$$\begin{aligned} S^{Fire} &= \sum_{i=1}^{N} R_i^{Fire}, \\ S^{MTPL} &= \sum_{i=1}^{M} R_i^{MTPL}. \end{aligned}$$

Let us define the aggregate liability of the ceding company for each line :

$$\begin{aligned} T^{Fire} &= \sum_{i=1}^{N} C_i^{Fire}, \\ T^{MTPL} &= \sum_{i=1}^{M} C_i^{MTPL}. \end{aligned}$$

Now let us assume that the ceding company buys a multiline excess of loss cover of the form

$$Cover = \max(0, S^{Fire} + S^{MTPL} - GAAD)$$

where $GAAD$ is a global annual aggregate deductible playing on both lines of business.
In this paper we are interested in analysing the retention's risk of the ceding company :

$$Retention = T^{Fire} + T^{MTPL} + \min(S^{Fire} + S^{MTPL}, GAAD).$$

2. Dependencies generated by the model

Analysing and modelling dependencies is a subject that received great attention during the last few years. Different methods have been proposed to tackle that problem, e.g. the use of Fréchet bounds (see e.g. Dhaene et al. (2001)) or the use of copulas (see e.g. Frees and Valdez (1998). These methods do not recognize the exact dependency structure because it is often not possible to model it.
In our case, there is clearly some dependency which does not allow an easy analysis of the problem. However the dependency in our model is induced by the model itself. We then have the chance to model the dependency exactly and possibly obtain exact calculations.
The fact that reinsurance induces dependencies has been observed by Walhin and Paris (2000) for the analysis of the cedent's retention's risk when there are paid reinstatements, by Walhin and Denuit (2003) for the practical pricing of Top & Drop covers, by Walhin (2003) for the pricing of exotic excess of loss covers. The present paper shows another dependency induced by the model.
Fortunately it is easy to make a modelization of our dependency : the random variables R_i^{Fire}, C_i^{Fire} depend on X_i^{Fire} whereas R_i^{MTPL}, C_i^{MTPL} depend on X_i^{MTPL}. This means that even though N, M, X^{Fire}, and Y^{Fire} are mutually independent, S^{Fire}, S^{MTPL}, T^{Fire}, T^{MTPL} are not which makes the calculation of the distribution of *Retention* difficult. We need to obtain the joint distribution of

$$(S^{Fire}, S^{MTPL}, T^{Fire}, T^{MTPL}).$$

In fact if we obtain the joint distributions of (S^{Fire}, T^{Fire}) and (S^{MTPL}, T^{MTPL}), we have a solution to our problem because these random vectors are independent thanks to the mutual independence hypotheses we made.
An easy solution is available and is described in the next section.

3. Bivariate Panjer's algorithm

Our problem fits exactly within the framework of the multivariate Panjer's algorithm, described in Walhin and Paris (2000), or in Sundt (1999).
We just need the bivariate setting in order to obtain the joint distributions we need.
Let us define :

$$\begin{aligned} f^i(x,y) &= \mathbb{P}[R^i = x, C^i = y] \quad , \quad i = Fire, MTPL, \\ g^i(s,t) &= \mathbb{P}[S^i = s, T^i = t] \quad , \quad i = Fire, MTPL. \end{aligned}$$

From now on we will not use the superscript anymore.
Let us assume that N belongs to the Panjer's family of counting distributions :

$$\frac{\mathbb{P}[N=n]}{\mathbb{P}[N=n-1]} = a + \frac{b}{n} \quad , \quad n \geq 1.$$

We have :

$$\begin{aligned} g(0,0) &= \Psi_N(f(0,0)), \\ g(s,t) &= \frac{1}{(1-af(0,0))} \sum_{x,y}^{s,t} [a + b\frac{x}{s}] g(s-x, t-y) f(x,y) \quad , \quad s \geq 1, \\ g(s,t) &= \frac{1}{(1-af(0,0))} \sum_{x,y}^{s,t} [a + b\frac{y}{t}] g(s-x, t-y) f(x,y) \quad , \quad t \geq 1, \end{aligned}$$

where

$$\begin{aligned} \sum_{x,y}^{s,t} g(x,y) &= \sum_{x=0}^{\min(s,m)} \sum_{y=0}^{\min(t,n)} g(x,y) - g(0,0), \\ m &= \max(x | f(x,y) > 0), \\ n &= \max(y | f(x,y) > 0). \end{aligned}$$

and $\Psi_N(u)$ denotes the probability generating function of N : $\Psi_N(u) = \mathbb{E}[u^N]$.

It is clear that the above-mentioned algorithm is time-consuming. However we will take advantage of the specific dependence structure in order to minimize the computing time. Indeed the random vector (R, C) has positive masses only along an S-shape. So we may

adapt the formula as :

$$
\begin{aligned}
g(0,0) &= \Psi_N(f(0,0)),\\
g(0,t) &= 0 \quad , \quad t \geq 1,\\
g(s,0) &= \frac{1}{(1-af(0,0))} \sum_{x=1}^{s} [a + b\frac{x}{s}] g(s-x,0) f(x,0) \quad , \quad 1 \leq s \leq D,\\
g(s,t) &= \frac{1}{(1-af(0,0))} \times \\
&\quad \left[\sum_{x=1}^{s} [a + b\frac{x}{s}] g(s-x,0) f(x,0) + \sum_{y=1}^{t} [a + b\frac{D}{s}] g(s-D,t-y) f(D,y) \right] \quad ,\\
&\quad 1 \leq s \leq D \quad , \quad 1 \leq t \leq L,\\
g(s,t) &= \frac{1}{(1-af(0,0)))} \times \\
&\quad \left[\sum_{x=1}^{s} [a + b\frac{x}{s}] g(s-x,0) f(x,0) + \sum_{y=1}^{t} [a + b\frac{D}{s}] g(s-D,t-y) f(D,y) + \right.\\
&\quad \left. \sum_{x=D+1}^{s} [a + b\frac{x}{s}] g(s-x,t-L) f(x,L) \right] \quad , \quad s > D \quad , \quad t \geq L.
\end{aligned}
$$

4. NUMERICAL APPLICATION

Let us make the following hypotheses for our numerical example :

- the distribution of the fire claim amounts, X^{Fire}, is limited Pareto with parameters $A = 400$, $B = 2000$ and $\alpha = 1.50$. The distribution of the MTPL claim amounts, X^{MTPL} is limited Pareto with parameters $A = 700$, $B = 2000$ and $\alpha = 2.50$. Let us recall the cumulative density distribution of a limited Pareto distribution ($X \sim Pa(A,B,\alpha)$) :

$$
\begin{aligned}
F_X(x) &= 0 \quad \text{if } x \leq A,\\
&= \frac{A^{-\alpha} - x^{-\alpha}}{A^{-\alpha} - B^{-\alpha}} \quad \text{if } A < x \leq B,\\
&= 1 \quad , \quad x > B.
\end{aligned}
$$

- the distribution of the fire claim numbers, N is Poisson with parameter $\lambda = 2.5$. The distribution of the MTPL claim numbers, M is Poisson with parameter $\lambda = 5$. Let us recall the probability function of a Poisson distribution ($N \sim Po(\lambda)$) :

$$
\mathbb{P}[N = n] = p(n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad , \quad n = 0, 1, \ldots
$$

Working with Poisson distributions allows us to work with the bivariate Panjer's algorithm. Moreover, it simplifies the use of the algorithm as $a = 0$ in the Poisson case.

We note that the limited Pareto distribution is a continuous distribution whereas we need a discrete distribution in order to use the bivariate Panjer's algorithm. We therefore choose to obtain a discretization of our limited Pareto distributions by using the local moment matching method with one moment (see Gerber (1982)). It is not difficult to show that the discrete version of a limited Pareto distribution is given by

$$\begin{aligned} f_{X_{dis}}(A) &= 1 - \frac{\frac{(A+h)^{1-\alpha}}{1-\alpha} - \frac{A^{1-\alpha}}{1-\alpha} - B^{-\alpha}h}{h(A^{-\alpha} - B^{-\alpha})}, \\ f_{X_{dis}}(A+jh) &= \frac{2(A+jh)^{1-\alpha} - (A+(j-1)h)^{1-\alpha} - (A+(j+1)h)^{1-\alpha}}{h(1-\alpha)(A^{-\alpha} - B^{-\alpha})} \\ & \quad j = 1, \ldots, \frac{B-A}{h} - 1, \\ f_{X_{dis}}(B) &= 1 - f_{X_{dis}}(A) - f_{X_{dis}}(A+h) - \cdots - f_{X_{dis}}(B-h). \end{aligned}$$

where h is chosen such that $\frac{B-A}{h}$ is an integer.
Obtaining the expected retained loss is easily given by

$$\mathbb{E}Retention = \mathbb{E}T^{Fire} + \mathbb{E}T^{MTPL} + \mathbb{E}\min(S^{Fire} + S^{MTPL}, GAAD).$$

As S^{Fire} and S^{MTPL} are independent, we do not need to apply the bivariate Panjer's algorithm. However as we will compute standard deviation and Wang Transforms of *Retention*, we will need the distribution of *Retention* and thus we will have to apply the bivariate Panjer's algorithm. We will also make the calculations with the false assumption of independence between S^{Fire} and T^{Fire} on the one hand and S^{MTPL} and T^{MTPL} on the other hand.
Our aim is now to analyse different reinsurance structures in order to find optimal reinsurance agreements. We will therefore let the deductibles and limits vary as well as the global annual aggregate deductible.
For each situation we are going to compute the following elements :

1. $\mathbb{E}Retention$
2. $\sigma(Retention)$
3. $WT_{0.90}(Retention)$
4. $WT_{0.95}(Retention)$
5. $WT_{0.99}(Retention)$

where $WT_{1-\alpha}(Retention)$ denotes the Wang Transform of level α of the random variable *Retention* (see Wang (2002)). Let us define

1. F the cumulative density function of the random variable *Retention*
2. $\Phi(.)$ the cumulative density function of the standard normal distribution
3. α a security level
4. $\lambda = \Phi^{-1}(\alpha)$

5. $F^*(x) = \Phi[\Phi^{-1}(F(x)) - \lambda]$

Then the Wang Transform of level α is given by the expectation of *Retention* under the measure F^* :

$$WT_{1-\alpha}(Retention) = \mathbb{E}^*(Retention).$$

A good situation for the insurer is when $\mathbb{E}Retention$ is as high as possible (in such a case, it means that the cession to the reinsurer is small which means in other words that the smallest expected profit is ceded to the reinsurer) and when the risk measure (either the standard deviation or the Wang Transform) is as small as possible (which means that few capital has to be allocated).

Let us first analyse the following case, which we denote Treaty 1 :

$$\begin{aligned} D^{Fire} &= 500, \\ L^{Fire} &= 1500, \\ D^{MTPL} &= 800, \\ L^{MTPL} &= 1200, \\ GAAD &= 0. \end{aligned}$$

Table 1: Treaty 1

We obtain the following quantities of interest :

$$\begin{aligned} \mathbb{E}(Retention) &= 3949.617, \\ \sigma(Retention) &= 1655.303, \\ WT_{0.90}(Retention) &= 6252.296, \\ WT_{0.95}(Retention) &= 6971.925, \\ WT_{0.99}(Retention) &= 8394.352. \end{aligned}$$

Table 2: Retained risk for Treaty 1

Assume that the ceding company does not agree with such a large cession. Then a natural solution is to increase the priorities of the treaties. We then move to Treaty 2 :

$$
\begin{aligned}
D^{Fire} &= 800, \\
L^{Fire} &= 1200, \\
D^{MTPL} &= 1000, \\
L^{MTPL} &= 1000, \\
GAAD &= 0.
\end{aligned}
$$

Table 3: Treaty 2

We obtain the following quantities of interest :

$$
\begin{aligned}
\mathbb{E}(Retention) &= 4642.687 \\
\sigma(Retention) &= 1949.410 \\
WT_{0.90}(Retention) &= 7355.088 \\
WT_{0.95}(Retention) &= 8202.904 \\
WT_{0.99}(Retention) &= 9878.696
\end{aligned}
$$

Table 4: Retained risk for Treaty 2

Obviously the objective is attained : the cession is now smaller. However, on the other hand the risk level is higher (larger standard deviation and larger Wang Transforms). This behaviour is obvious. Now let us move to Treaty 3 which is the same than Treaty 1 but with a global annual aggregate deductible :

$$
\begin{aligned}
D^{Fire} &= 500, \\
L^{Fire} &= 1500, \\
D^{MTPL} &= 800, \\
L^{MTPL} &= 1200, \\
GAAD &= 1000.
\end{aligned}
$$

Table 5: Treaty 3

We obtain the following quantities of interest :

$\mathbb{E}(Retention)$	=	4756.575
$\sigma(Retention)$	=	1822.765
$WT_{0.90}(Retention)$	=	7202.147
$WT_{0.95}(Retention)$	=	7939.854
$WT_{0.99}(Retention)$	=	9381.442

Table 6: Retained risk for Treaty 3

We immediately observe that this treaty is optimal with respect to Treaty 2 : the cession is smaller and the retained risk is also smaller. So clearly Treaty 3 is a better choice than Treaty 2.
Other situations may be described. For example, let us compare Treaty 4 with Treaty 5 :

	Treaty 4	Treaty 5
D^{Fire}	1000	500
L^{Fire}	1000	1500
D^{MTPL}	1200	800
L^{MTPL}	800	1200
$GAAD$	0	2000

Table 7: Treaties 4 and 5

We obtain the following quantities of interest :

	Treaty 4	Treaty 5
$\mathbb{E}(Retention)$	4946.616	5150.214
$\sigma(Retention)$	2103.647	2093.537
$WT_{0.90}(Retention)$	7884.110	7921.404
$WT_{0.95}(Retention)$	8804.185	8729.225
$WT_{0.99}(Retention)$	10626.00	10266.98

Table 8: Retained risk for Treaties 4 and 5

Here again, we observe that Treaty 5 is optimal with respect with Treaty 4 : smaller cession and smaller retained risk.
Now let us compute the quantities of interest of Treaty 5 with the wrong assumption of independence. We obtain :

$$
\begin{aligned}
\mathbb{E}(Retention) &= 5150.214 \\
\sigma(Retention) &= 1777.361 \\
WT_{0.90}(Retention) &= 7584.320 \\
WT_{0.95}(Retention) &= 8332.368 \\
WT_{0.99}(Retention) &= 9800.117
\end{aligned}
$$

Table 9: Retained risk for Treaty 5 with wrong assumption of independence

As explained above, the expected retention is the same as in the exact model. However the risk measures are smaller in the wrong model which is logical because the wrong model ignores the positive dependence that is present in the model.
Using the standard deviation of the $WT_{0.99}$ as the risk criterion, we may conclude that Treaty 5 is optimal with respect to Treaty 2. In fact, using the exact model, we immediately see that we are not allowed to give such a conclusion. This shows the danger of working with the model ignoring the dependencies.

5. Conclusion

We have analysed an actuarial situation where dependence is induced by the model. This kind of dependence was tractable by using the multivariate Panjer's algorithm. We have been able to show, on our numerical example the danger of working within a wrongly assumed model where there is no dependence and we also have shown the optimality of the multiline excess of loss cover. Some practical considerations are

1. It may be the case that the loading of the insurer and reinsurer are very different. Then the optimality should be studied with respect to the expected gain and not with respect to the expected retention.

2. We here have analysed the large claims, that are reinsured through an excess of loss treaty. Obviously we should account for the small claims in order to compute the risk measures.

3. This paper says that it is better for the ceding company to buy excess of loss treaties with small priorities and with a global annual aggregate deductible. Administrative reasons may go against these solutions. Indeed, small priorities means that large number of claims are expected to hit the layers, which makes lots of administration for both the insurer and the reinsurer. This is in particular true for long-tailed business like MTPL where a stability clause is generally in use.

4. In practice, the reinsurer would limit its annual liability through a global annual aggregate limit.

5. When the priorities of the treaties tend to 0 and the limits tend to infinity, then the cover becomes a multiline stop-loss treaty.

The hypothesis of independence between the lines of business may be relaxed for the case of umbrella covers where correlations exist between the covered lines of business. In such a case, copulas may help in order to price the cover. However, in order to analyse the retained risk of the cedant, only simulations would help and one should be aware of the fact that a huge number of simulations would be necessary in order to correctly catch the dependencies in the tails.

REFERENCES

Frees, E. and Valdez, E.A. (1998), Understanding Relationships Using Copulas, North American Actuarial Journal, 2, 1-25

Gerber, H.U. (1982), On the Numerical Evaluation of the Distribution of Aggregate Claims and its Stop-Loss Premiums, Insurance:Mathematics and Economics, 1, 13-18

Panjer, H.H. (1980), Recursive Evaluation of a Family of Compound Distributions, Astin Bulletin, 12, 22-26

Ribeaud, D. (2000), Multiline-Rückversicherungslösungen, Swiss Reinsurance Company, Zurich

Sundt, B. (1999), On Multivariate Panjer's Recursions, Astin Bulletin, 29, 29-46

Walhin, J.F. and Paris, J. (2000), The Effect of Excess of Loss Reinsurance with Reinstatements on the Cedant's Portfolio, Blätter der Deutsche Gesellschaft für Versicherungsmathematik, 24, 615-627

Walhin, J.F. and Denuit, M. (2003), On the Pricing of Top and Drop Excess of Loss Covers, submitted

Walhin, J.F. (2002), On the Practical Multiline Excess of Loss Pricing, CAS Forum, Winter 2001, 121-151

Walhin, J.F. (2003), Some Comments on the Pricing of Exotic Excess of Loss Treaties, to be published in Journal of Actuarial Practice

Wang, S. (2002), A Risk Measure that Goes beyond Coherence, discussion paper Waterloo University

Hedging Catastrophe Risk Using Index-Based Reinsurance Instruments

Lixin Zeng

Hedging Catastrophe Risk Using Index-Based Reinsurance Instruments

Lixin Zeng

March 2003

Abstract

Index-based hedging instruments such as industry loss warranties are increasingly recognized as effective hedging tools for insurance and reinsurance portfolios. However, wider adoption of these instruments is inhibited by basis risk, the difference between the index-based payoff and the buyer's actual loss. This study presents a systematic approach for potential buyers to analyze and manage basis risk in order to take full advantage of the benefits offered by these instruments.

We examine two measures of basis risk: (i) hedging effectiveness and (ii) conditional payoff shortfall. Many existing measures such as hedge volatility and correlation are special cases of the hedging effectiveness measure. Next, we study the tradeoff between basis risk and the cost of hedging. Finally, we present a robust numerical algorithm designed to optimize an index-based hedging program consisting of multiple index-based contracts.

1. Introduction

In recent years, we have observed growing interest in index-based hedging instruments, especially in the areas of catastrophe risk reinsurance and securitization. Examples include industry loss warranty (ILW) contracts and index-based cat bonds. In contrast to a traditional indemnity-based reinsurance contract, an index-based instrument has a payoff that is not completely determined by the loss incurred by the purchaser[1]. Instead, it is determined by an index that is positively correlated with the purchaser's actual loss. The index can be the industry loss or certain meteorological or seismic parameters related to a natural disaster event. The most frequently used industry loss indices used in the US are based on incurred insurance losses surveyed and published by the Property Claims Service.

The main advantage of index-based instruments is that they are practically free from moral hazard, a major hurdle that discourages capital market investors from participating in insurance risk securitization, even though the natural catastrophe risk is an extremely appealing asset class from a portfolio perspective (Litzenberger *et. al.*, 1996). The absence of moral hazard also suggests that an index-based instrument should command a lower margin than a comparable indemnity-based reinsurance contract (Cummings, *et. al.*, 2003), making it an attractive alternative to traditional reinsurance. Moreover, it is shown in Doherty and Richter (2002) that combining indemnity contracts with index-based instruments can ideally lead to efficiency gains for purchasers.

However, index-based instruments pose a new challenge to the purchasers in the form of basis risk - the difference between the actual loss experienced by the purchaser and the payoff of the index-based contract. The difference is one of the primary factors that have kept many potential purchasers away from these instruments. A systematic, credible, and practical way to quantify and manage basis risk must be made available to the potential

[1] Currently, the purchasers of index-based instruments are almost exclusively insurance and reinsurance companies. However, end users of insurance (e.g., corporations) have started exploring the use of this type of instruments.

purchasers before index-based instruments can gain recognition as a main stream risk management tool and become widely adopted.

The task of quantifying and managing basis risk can be divided into two problems: First, given an existing portfolio of liabilities to be hedged and an index-based hedging program consisting of one or more index-based contracts, how best to quantify the basis risk associated with this hedging strategy? Second, given an underlying portfolio and a set of constraints reflecting the buyer's risk appetite and return requirement, how can one construct an index-based hedging program to achieve an optimal balance between cost and hedging effectiveness?

This study focuses on these two issues. In Section 2, we state the assumptions and notations used in this study. Next, we develop an analytical framework to quantify basis risk in an effort to unify commonly used measures of basis risk (Section 3). In Section 4, we introduce an approach to construct an index-based hedging program that optimally balances hedging effectiveness and cost while satisfying certain constraints. Section 5 summarizes the study.

2. Assumptions and notations

We do not assume any specific form of parametric distribution for the random variables such as losses and underwriting profits. Instead, we represent the randomness of the "state of the world" using a large number of scenarios. This is because our primary interest is in hedging catastrophe risk and the outputs of most catastrophe models, which serve as inputs to our analyses, are scenario-based. In addition, although the numerical examples presented in this paper are realistic, they are hypothetical and are not based on any specific catastrophe model or actual company data.

Furthermore, we make three simplifications. First, it is assumed that only one loss event occurs in a year, although the analyses presented can be extended to include multiple events on an annual aggregate basis without difficulty using existing dynamic financial analysis (DFA) tools. However, not including DFA allows us to simplify the equations and focus on basis risk analysis. For the same reason, we also ignore premium

reinstatement provisions frequently observed in actual transactions. Second, we do not consider the potential basis risk arising from the counterparty credit risk (i.e. the risk that the seller of the hedging contract fails to fully perform its contractual obligation). This permits us to focus on the discrepancy caused by the general lack of a one-to-one relationship between the actual loss and the index value. Third, we use binary ILW contracts in all examples. Nevertheless, the methodology developed can be applied to other forms of index-based instrument without substantial modification.

Lower and upper case letters are used to represent deterministic and random variables, respectively. Let L be the actual loss and X be the payoff of a hedging instrument. X is a function of an index I:

$$X = g(I) \tag{1}$$

For a binary ILW, I is the predefined industry loss for a region and given peril(s), and the payoff is defined as

$$X_I = g_I(I) = \begin{cases} l, I \geq i_t \\ 0, I < i_t \end{cases} \tag{1a}$$

where l is the limit of the ILW and i_t is known as the *trigger* of the contract. Another special case is an indemnity reinsurance policy, where $I = L$ and the payoff is defined as

$$X_R = g_R(L) = \begin{cases} l', L \geq r + l' \\ 0, L < r \\ L - r, r \leq L < l' \end{cases} \tag{1b}$$

where r and l' are the retention and limit of the reinsurance policy, respectively.

The net post-hedging loss L^* is then.

$$L^* = L - X = L - g(I) \tag{2}$$

It is possible that X > L. However, from an accounting point of view, this will not be allowed if the buyer wishes to treat the hedging instrument as reinsurance. Hence, Equation 2 frequently takes the following form

$$L^* = \max[0, L - X] = \max[0, L - g(I)] \tag{2a}$$

which forbids the buyer from claiming more than the actual loss. Specifically, we use L^*_I and L^*_R to denote the net loss after an ILW and an indemnity reinsurance policy, respectively:

$$\begin{aligned} L^*_I &= \max[0, L - g_I(I)] \\ L^*_R &= L - g_R(L) \end{aligned} \tag{2b}$$

3. Definition and quantification of basis risk

3.1. The cause of basis risk – a qualitative view

With an indemnity reinsurance policy, the amount of payoff is always precisely predictable given an actual loss, even though the actual loss itself is random (e.g., Equation 1b). However, this is generally not true for index-based instruments. We consider a hypothetical insurer (Company A), which has a geographically diversified exposure in the region where it sells property insurance and is considering using an ILW to hedge its catastrophe risk. As shown in Figure 1, at a given level of actual loss (e.g., along the dashed horizontal line), the industry loss index cannot be uniquely determined *a priori*. As a result, if Company A buys an ILW (Equation 1a) with a trigger represented by the vertical dashed line in Figure 1, the ILW payoff can be either zero or *l*, represented by the scenarios to the left and the right of the vertical line, respectively. This randomness makes it impossible for a buyer to precisely predict the payoff as a function of the actual loss. Next, we attempt to quantify such randomness, which is known as the "basis risk".

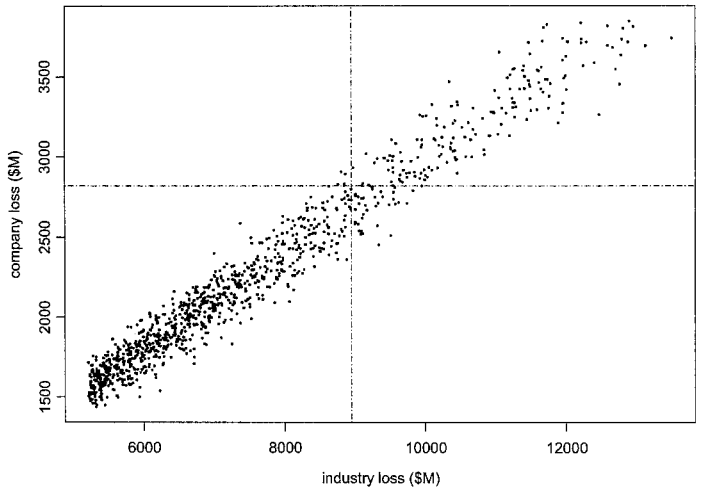


Figure 1. The loss to Company A vs. the industry loss. Each point in the plot represents a loss scenario. The dashed horizontal and vertical lines represent given levels of company and industry losses, respectively.

3.2. Benchmarks for comparison

Although basis risk is caused by the random difference between the index-based payoff (X_I) and actual loss (L), it is not sensible to directly compare X_I and L because rarely does a buyer expect the actual loss to be fully hedged. In the context of hedging catastrophe risk, the focus of the buyer is on reducing the severity of *large* losses. Hence, it is more meaningful to compare X_I to the payoff of a benchmark indemnity reinsurance policy (X_R) or, equivalently, compare the net losses associated with the index-based instruments and the benchmark, i.e. L^*_I vs. L^*_R (Cummings, *et. al.*, 2003).

The choice of the benchmark is usually based on the risk management objective of the buyer. For example, Company A currently has an annual probability of default[2] of 1%; a

[2] For illustration purpose here, the company is considered in default if the loss exceeds its surplus.

change in business environment requires this probability to be reduced to 0.4%. The traditional reinsurance approach to accomplish this is to purchase an indemnity reinsurance policy with the retention $r = v_0$ and the limit $l' = v_1 - v_0$, where v_1 and v_0 are the 99th and 99.6th percentile value at risk (VaR) of the underlying portfolio. Hence, its payoff (X_R) and net loss after this reinsurance (L^*_R) can serve as the respective benchmarks for the payoff (X_I) and net loss after an ILW (L^*_I). The cumulative distribution function of the loss of the underlying portfolio is shown in Panel I of Figure 2.

Next, we attempt to use an ILW to accomplish the same objective stated above. We choose the 99th percentile of the industry loss as the trigger and $v_1 - v_0$ as the limit, denoted i_t and l, respectively (Equation 1a). The basis risk of the ILW can then be defined based on the difference between L^*_I and L^*_R.

3.3. Definition and quantification of basis risk

The cumulative distribution functions (CDF) of L^*_R and L^*_I are shown in Panels II and III of Figure 2. Since L^*_R and L^*_I are random, we can compare their respective statistical summaries or evaluate the statistical summaries of their difference ($L^*_R - L^*_I$). These comparisons lead to the definitions of two types of basis risk.

Basis Risk Related to Hedging Effectiveness (Type I): In general, the purpose of purchasing a hedging instrument (reinsurance or ILW) is to reduce the risk of the underlying portfolio. The hedging effectiveness of the instrument can be measured by the amount of risk reduced. Let h_r and h_i denote the hedging effectiveness of the benchmark and the ILW. They can be defined as

$$\begin{aligned} h_r &\equiv 1 - y_r / y_g \\ h_i &\equiv 1 - y_i / y_g \end{aligned} \tag{3a}$$

where y_g, y_r, and y_i are the statistical measures of the risk of the underlying portfolios before any hedging, net of the benchmark, and net of the ILW, respectively. Frequently used risk measures include standard deviation, value at risk (VaR), tail value at risk (TVaR), and probability of default (POD). The choice of the proper risk measure has

been extensively discussed in the actuarial literature (e.g., Artzner *et. al.*, 1999) and is not repeated here.

The Type I basis risk (referred to as b_I hereafter) measures the hedging effectiveness of an index-based instrument relative to that of the benchmark. Hence, it can be defined as

$$b_I \equiv 1 - h_i / h_r \tag{3b}$$

where we assume the benchmark hedging always reduces risk, i.e. $h_r > 0$.

Equation (3b) is obviously not the only valid definition. In fact, any b_I that increases with decreasing h_i/h_r is a valid quantification of basis risk. Partially due to this reason, basis risk is not uniquely defined in previous studies. For example, Major (1999) uses volatility of hedging to represent basis risk, whereas Harrington and Niehaus (1999) and Meyers (1996) measure basis risk based on the linear correlation coefficient between the actual loss and index-based payoff.

For Company A, the selected risk measure is the probability of default (POD), as reducing POD is its objective of hedging. Since the ILW does not reduce POD to the desired benchmark level, a substantial amount of basis risk exists (Table 1).

Table 1. Numerical values of hedging effectiveness and basis risk related to the ILW structure defined in Section 3 for Company A

	Underlying portfolio	Net of indemnity reinsurance	Net of ILW
Probability of default (risk measure)	1.00%	0.40%	0.60%
Hedging effectiveness		60.0%	40.0%
b_I			33.3%

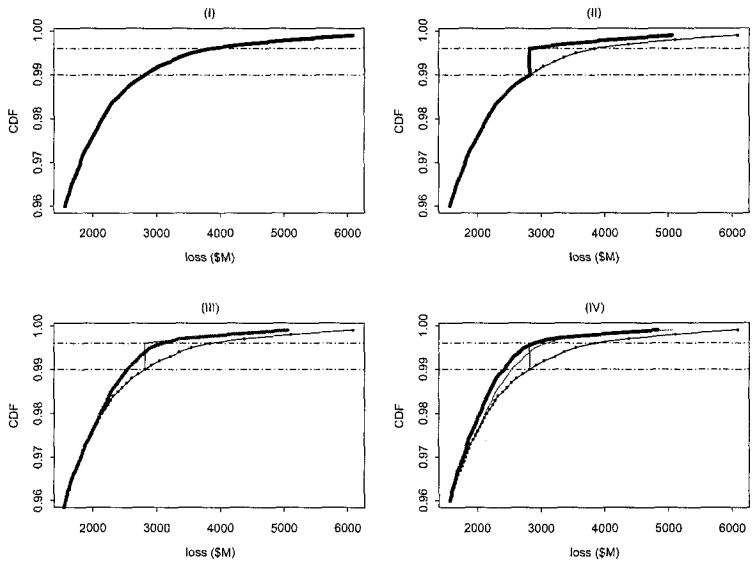


Figure 2. The cumulative distribution functions (CDF) of the gross and net losses of the underlying portfolio of Company A: (I) without hedge; (II) the thick curve: net of the benchmark; (III) the thick curve: net of the ILW defined in Section 3.2; (IV) the thick curve: net of the optimal ILW defined in Section 4.2. The thin curves in each of the panels (II), (III) and (IV) are the same curves as those in the previous panels for comparison purposes. The two horizontal dashed lines represent the 99% and 99.6% quantiles of the CDF.

Basis Risk of Payoff Shortfall (Type II): In general, two hedging instruments that accomplish the same hedging effectiveness do not guarantee the same payoff. Hence, even if b_I for an index-based hedging instrument is zero, it is still possible that the index-based payoff is less than the benchmark. To account for such discrepancy, we define the payoff differential (ΔL^*) as:

$$\Delta L^* \equiv X_I - X_R = L_R^* - L_I^* \tag{4}$$

where X_I, X_R, L^*_R, and L^*_I, defined in Section 2, are the index-based payoff, reinsurance payoff, net loss after the benchmark reinsurance, and net loss after the index-based product, respectively. A negative value of ΔL^* indicates that the buyer of the index-based instrument would recover more if the benchmark indemnity instrument were used instead (i.e. there is a payoff shortfall for the index-based instrument). This is another important aspect of basis risk in addition to its impact on hedging effectiveness. Because the purchaser is generally interested in protection against large losses, we examine the conditional cumulative distribution function of ΔL^* given the occurrence of a loss severe enough to trigger the payoff of the benchmark (i.e. $X_R>0$). The *conditional CDF* is simply denoted as $f_b(s)$:

$$f_b(s) \equiv prob(\Delta L^* < s \mid X_R > 0) \tag{5}$$

where $prob(\bullet)$ stands for the probability that "•" occurs. Examples of $f_b(s)$ are shown in Figure 3. Since we are primarily interested in measuring the downside risk of index-based instruments, we define the *Type II Basis Risk* (referred to as b_2 hereafter) as:

$$b_2 \equiv \frac{\max(-s^\alpha, 0)}{l'} \tag{6}$$

where s^α is the α^{th} quantile of $f_b(s)$. Under this definition, b_2 is the quantile of the index-based payoff shortfall normalized by the limit of the benchmark indemnity reinsurance policy (l'). For the ILW structure defined above for Company A, selected values of b_2 are listed in Table 2. The last row in the table shows that, for example, *given* the occurrence of a loss greater than the benchmark retention (r), there is a probability of 0.05 that the index-based payoff shortfall will exceed 19.9% of the limit of the benchmark hedging program.

Table 2. Selected values of b_2 for the initial ILW structure defined in Section 3.2 and the optimal ILW defined in Section 4.2 for Company A.

α	b_2 *(initial)*	b_2 *(optimal)*
0.004	43.4%	19.3%
0.01	41.1%	17.7%
0.05	19.9%	1.8%

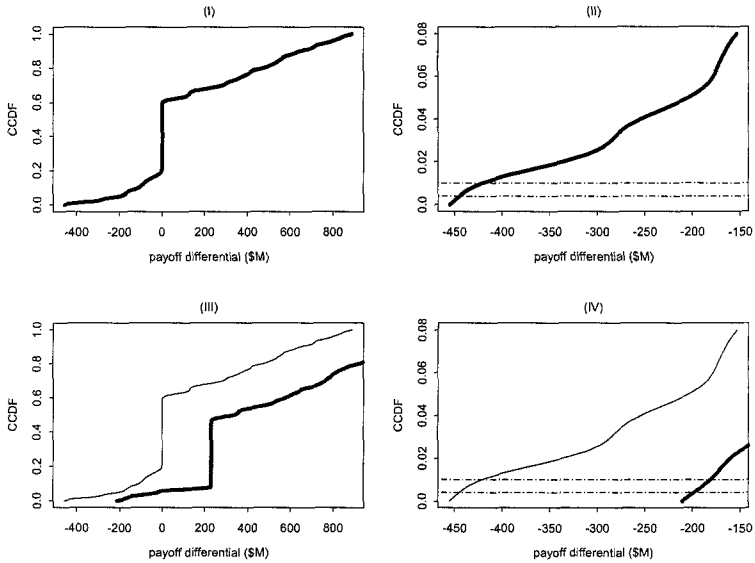


Figure 3. the conditional CDF of payoff differentials. Panels I and II: for the ILW structure defined in Section 3; the negative tail of the curve in I is a shown in II. Panels III and IV: for the ILW structure defined in Section 3.2 (thin lines) and for the optimal ILW defined in Section 4.2 (thick lines); the negative tail of the curve in III is a shown in IV. The horizontal dashed lines in Panels II and IV represent the 0.4% and 1% quantiles of the conditional CDF.

In summary, b_1 measures the hedging effectiveness of an index-based instrument relative to a benchmark, which is usually an indemnity reinsurance policy. Because this is directly related to the risk/return profile of the net post-hedge portfolio, we believe b_1 should be the focus of the buyer in evaluating the benefit of index-based strategies. However, b_2 is also important in practical decision-making as it measures the "probability of regret" for choosing an index-based instrument over a more traditional indemnity reinsurance policy. In this context, b_2 does not reflect or give any value to the fortuitous

gain[3] available from the index-based instrument, which must be taken into account for the purpose of designing an optimal index-based hedging program (Section 4).

4. Optimizing an index-based hedging program

4.1. An overview

When the basis risk associated with an index-based instrument exceeds a tolerable threshold established by the purchaser, the contract terms must be modified such that the basis risk is reduced to the acceptable level. Given an underlying portfolio, there are primarily two ways to accomplish this: (a) changing the index or indices used by the contract and/or (b) modifying the parameters associated with each index (e.g., trigger and limit). It is possible that the cost of the contract will increase due to these changes. An optimal contract best balances the cost and benefit while satisfying the constraints imposed on the buyer. The process of arriving at such an optimal balance is illustrated using a simple example (Section 4.2). A robust method for optimizing complicated real world index-based contracts is introduced in Section 4.3.

4.2. A simple example

We revisit the example of Company A. We assume that the company wishes to reduce the basis risk associated with the initial ILW defined in Section 3.2 by changing the limit and trigger of the ILW. Specifically, it wishes to accomplish the following two objectives:

(a) Reduce b_I to zero (i.e. it requires that the ILW has the same level of hedging effectiveness as the benchmark). In this case, the task is to reduce the POD net of the ILW from 0.6% to 0.4%.

(b) Achieve Objective (a) with the lowest possible cost, allowing the underlying portfolio to retain the maximum possible net expected profit.

[3] the fortuitous gain is referred to as the excess recover from an index-based instrument relative to the benchmark (i.e. when $\Delta L^* > 0$). Under reinsurance accounting, it is impossible for the buyer to recover more than its gross pre-hedging loss.

Hence, the optimal ILW in this case is one that maximizes the net expected profit of the underlying portfolio while keeping POD from exceeding 0.4%.

We first plot how POD varies as a function of the ILW trigger and limit (the thick contours in Figure 4). The POD is represented by the contours of equal POD values. A point on a contour labeled x represents the trigger/limit combination of an ILW contract net of which the underlying portfolio has a POD of x. We call such a contour the *equal POD curve of* x (e.g., 0.4%). All points located to the upper-left of the curve correspond to POD less than x, and vice versa.

The initial ILW is represented by the solid square, which is located on the equal POD curve of 0.6%. For the POD to be reduced to 0.4% or less, the limit and trigger combination must be adjusted such that it is located on or to the upper-left of the equal POD curve of 0.4%. In fact, an ILW represented by any point on the equal-POD line of 0.4% can achieve the first objective.

We next examine the costs associated with different ILW contracts in order to accomplish the second objective. It is assumed that the premium for the contract is equal to five times the expected payoff, representing a typical profit margin of this type of contract in the market. With this assumption, the net expected profit[4] is calculated and visualized as the thin contours in Figure 4. A point on a contour labeled y represents the trigger/limit combination of an ILW contract, net of which the underlying portfolio has an expected profit of y. We simply call such a contour the *equal profit curve of* y. All points located to the upper-left of the curve correspond to net expected profits less than y, and vice versa.

[4] Defined as the premium of the underlying portfolio minus the sum of (i) the cost of the ILW, (ii) the expected value of the net loss, and (iii) other expenses. These quantities are formally defined in Section 4.3

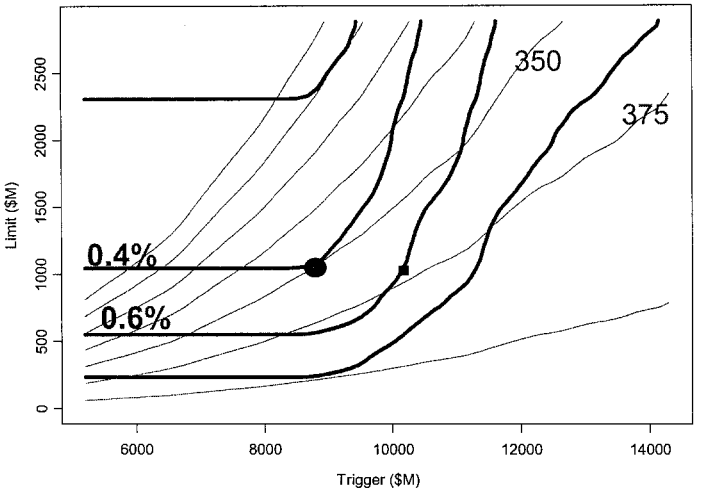


Figure 4. The probability of default (thick contour) and the expected profit (thin contour, in $M) net of ILW as a function of the trigger and limit. The solid square represents the initial ILW defined in Section 3.2, of which the index trigger is equal to the 100-year industry loss and the limit is equal to the difference between the buyer's 250-year loss and 100-year loss. The solid circle represents the ILW with the optimal trigger and limit arrived at in Section 4.2.

The point where the equal POD curve of 0.4% is tangent to an equal net profit curve is represented by the solid circle in Figure 4. The equal net profit curve represents a net expected profit of $350M. The solid circle represents the optimal ILW that accomplishes both objectives of the company because

(a) Since it is located on the equal POD curve of 0.4%, the first objective is achieved.

(b) All other points along and to the upper-left of the equal POD curve of 0.4% are also to the upper-left of the equal net profit curve of $350M. Hence, the

net profits associated with these points are less than that associated with the solid circle. Thus, the solid circle represents the trigger/limit combination corresponding to the greatest net expected profit, i.e. the combination that accomplishes the second objective.

By definition, b_1 is reduced to zero. The loss distribution function of the underlying portfolio net of the optimal ILW is shown in panel IV of Figure 2. b_2 is shown in panel IV of Figure 3 and in Table 2.

This simple example shows that, in general, the task of optimizing an index-based hedging program is essentially a problem of optimally balancing basis risk and costs. Once the buyer determines the amount of acceptable basis risk and, if any, other constraints, an optimal hedging program should maximize an objective function specified by the user. In the example above, the objective function is the net expected profit. Other commonly used objective functions include risk-adjusted return on capital, Sharpe Ratio, etc. (e.g., Zeng, 2000). The optimization problem is formalized and generalized in the next subsection.

4.3. A robust method for optimizing an index-based hedging program

A robust method for optimizing an index-based hedging program is needed to handle real world tasks primarily because the underlying portfolio frequently consists of exposures in multiple lines of business and geographical regions. Thus, the number of indices involved is usually significantly greater than one. This makes the exhaustive search method used above impractical. In addition, it is not feasible to vary the limit and trigger continuously to create an ideal contract because only ILWs available in the market can be purchased. In fact, we can control only the amount to purchase for each contract.

For the k^{th} contract available in the market ($k = 1, 2, \ldots, m$), where m is the number of different contracts available, we define the following

I_k the underlying index (e.g. industry loss index for a specific region);

z_k the amount purchased;

$\eta_k(I_k)$ the unit payoff function.

p_k the unit premium (i.e. cost per z_k).

The payoff and cost of contract k are $z_k\eta_k(I_k)$ and $z_k p_k$, respectively. They are partitioned into the product of the amount of contract purchased and their respective unit values because the amount z_k is a decision to be made by the optimization procedure whereas the unit values depends on the contract itself, regardless the amount purchased[5]. For a simple binary ILW, p_k and z_k are simply the *rate on line* and the limit purchased, respectively. The payoff (Equation 1a) can be rewritten as

$$\begin{aligned} \eta_k(I_k) &= \begin{cases} 1, I_k \geq i_t \\ 0, I_k < i_t \end{cases} \\ z_k &= l \\ g_I(I_k) &= z_k \eta_k(I_k) \end{aligned} \tag{7}$$

The total payoff (X) and total cost (p_t) of the hedging program are

$$\begin{aligned} X &= \sum_{k=1}^{m} z_k \eta_k(I_k) \\ p_t &= \sum_{k=1}^{m} z_k p_k \end{aligned} \tag{8}$$

Hence, the loss net of the hedging program (L^*, defined in Equation 2b) can be specifically rewritten as

$$L^* = \max[0, L - \sum_{k=1}^{m} z_k \eta_k(I_k)] \tag{9}$$

The expected profit prior to hedging (EP) and the expected profit net of hedging (EP^*) can be expressed as

$$\begin{aligned} EP &= q_0 - EL \\ EP^* &= q_0 - p_t - EL^* \end{aligned} \tag{10}$$

[5] The unit premium actually depends on the amount purchased due to the supply-demand balance; however, this dependency is not considered in the analyses to simplify the formulas.

where q_0 is the inward premium of the underlying portfolio net of expense[6]. E is the expected value operator on a random variable.

The goal of the optimization procedure is to find the set of $z = \{z_1, z_2, \ldots, z_m\}$ such that a general objective function φ is maximized and a series of constraints are satisfied. Most frequently used φ is the expected profit of the underlying portfolio scaled by a risk measure. For the example, it can be defined as

$$\varphi = \frac{EP^*}{y_i} \tag{11}$$

where EP^* is the net expected net profit and y_i is some measure of the risk of the portfolio net of hedging. The latter can be the standard deviation, value at risk, tail value at risk and/or other statistics of the net loss L^*. The constraints can be expressed as

$$\begin{aligned} &\psi_c(EP, p_t, y_i, b_1, b_2) \le 0 \\ &c = 1,2,\ldots,n_c \end{aligned} \tag{12}$$

where n_c is the number of constraints. The constraints usually reflect limitations on the overall risk of the portfolio and/or the total cost of hedging. It is possible that a constraint can completely satisfy the risk control need of the hedger; consequently, the objective function does not need to be scaled by a risk measure, as illustrated in the simple examples in Section 4.2. In this example, there is one single constraint requiring that the probability of default net of hedging (denoted POD^*) do not exceed 0.4%. The objective is to maximize the net profit subject to this constraint. The objective and constraint for this example can be expressed as:

$$\begin{aligned} &\varphi = EP^* \\ &\psi_1 = b_1 = POD^* - 0.4\% \le 0 \end{aligned} \tag{13}$$

In general, given concrete expressions of φ, ψ_c, and y_i, which are chosen by the buyer of the hedging program, all the independent variables in Equations 11 and 12 are functions

[6] If L only contains catastrophe losses computed by a cat model, then the expected non-cat loss should also be excluded from q_0.

of z only. Therefore, the values of φ and ψ_c are functions of z only. Then, the optimization task becomes searching for z such that $\varphi=\varphi(z)$ is maximized, subject to $\psi_c=\psi_c(z) \leq 0$, c=1,..., n_c.

If φ and ψ_c were linear or other smooth functions of z, this optimization task would be relatively easy to handle using traditional numerical algorithms such as the ones based on the steepest descent. However, because of the payoff function used in real-world transactions (e.g., Equation 7) are nonlinear and inherently not smooth, traditional optimization algorithms frequently fail to reach the global maximum.

In this study, an optimization procedure based on the genetic algorithm (GA) is used. Genetic algorithms are computing algorithms that simulate the mechanics of natural selection and natural genetics to "evolve" toward the optimal solution to problems. They are frequently applied to optimization problems where traditional approaches fail because of nonlinear, non-smooth, or discrete objective functions and constraints. A thorough discussion of GA is beyond the scope of this paper; however, interested readers can refer to, e.g., Goldberg (1989). The application of GA on index-based hedging is also introduced in Cummings, *et .al.* (2003). In this paper, we describe only the principle of this approach in the context of our task.

At first, randomly selected initial values are assigned to z to form the *original generation* (denoted z_0). Multiple individuals of the first generation ($z_{11}, z_{12},\ldots z_{1p}$) are created by randomly perturbing z_0, where p is the number of individuals; these p individuals are known as the *population* for this generation. A score for z_{1j}, based on the objective function and the constraint functions, is calculated to measure how "good" z_{1j} is. If any of the constraints is not satisfied, the score will be a large negative value (e.g., -10^{36}). If all constraints are satisfied, the score will be equal to $\varphi(z_{1j})$. The next generation population is created by combining two randomly selected individuals from the previous generation plus some random variations. The p individuals with the highest scores are retained (z_{21}, $z_{22},\ldots z_{2p}$). This process is repeated until a stopping condition is reached. For example, the stopping condition can be that the highest score among all populations in the current generation is very close to that in the previous generation. Upon stopping, the optimal z is

the z_{gj} (i.e. the j^{th} individual in the g^{th} generation) with the highest score (hence the greatest value of the objective function) among all individuals. It is the random combination of individuals that allows the optimization procedure to "escape" from local maximums and have a much better chance to reach the global maximum.

We illustrate this approach using the following example. Company B has an underlying portfolio with exposure in two regions. It uses the 99th percentile VaR to measure risk (y_i); its goal is to reduce y_i to a target level while maximize the objective function defined by Equation 11. The objective and constraint are listed in Table 3. The ILW contracts available in the market are summarized in Table 4.

Table 3. Objective and constraint of optimizing an index-based hedging program

	Inward premium ($K)	Expected annual loss ($K)	Expected profit ($K)	99th percentile VaR ($K)	φ
underlying portfolio	10,000	2,305	7,695	54,861	14%
objective of hedging				less than 30,000	maximize

Table 4. Price and availability of ILW contracts

region	trigger ($M)	rate-on-line (p_k)	Capacity available ($M)	amount purchased (**z**)
A	3,500	10%	20	z_1 = ?
A	10,000	6%	30	z_2 = ?
B	7,000	10%	25	z_3 = ?
B	20,000	6%	50	z_4 = ?

The task is to find the set of $\mathbf{z} = \{z_1, z_2, z_3, z_4\}$ such that, net of the hedging program, the objective stated above is accomplished. In addition, the market data above impose another constraint: the maximum value of $\{z_1, z_2, z_3, z_4\}$ cannot exceed their respective available capacities (i.e. maximum limits). This example represents real world problems closely except a very small number of available contracts is used ($m = 4$), which allows us to verify the results using exhaustive search (i.e. testing all possible combinations of z_1, z_2, z_3, and z_4). Nevertheless, the speed performance of this approach for larger m is shown to be acceptable. Table 5 outlines the compositions of the hedging program recommended by GA and exhaustive search. The risk and return statistics of the portfolio net of the hedging programs are summarized in Table 6. Although the objective function

considers only one risk measure (VaR), two additional commonly used risk measures (TVaR and standard deviation) are listed in the table for comparison.

Table 5. Optimal hedging program recommended by GA and exhaustive search. All values are in $K.

	z_1	z_2	z_3	z_4
Genetic algorithm	231	17222	24625	29563
Exhaustive search	0	17000	24500	29500

Table 6. Risk and return statistics of portfolio net of optimal hedging programs designed based on GA and exhaustive search. All values are in $K except the ratio φ

	Inward premium	Cost of hedging	Expected annual loss	Expected profit	99th percentile VaR	φ	99th percentile TVaR	Standard deviation
Underlying portfolio	10,000	-	2,305	7,695	54,861	14.0%	151,513	19,872
Net of optimal hedging – GA	10,000	5,270	1,312	3,419	14,419	23.7%	106,899	15,924
Net of optimal hedging – exhaustive search	10,000	5,240	1,317	3,443	14,641	23.5%	107,093	15,937

The GA-based results are very close to the benchmark solution produced by exhaustive search. In fact, it is better than the exhaustive search, in which the incremental value of z is only $500K. Although it is impossible to directly verify the results of the GA-based results using exhaustive search for larger m due to computational constraints, we believe that the GA-based algorithm remain accurate because it does not rely on any assumptions about m.

It is well known that most financial optimization procedures are subject to parameter risk, which can adversely affect the robustness of any optimal solution. For example, if TVaR is substituted for VaR as the risk measure in the objective function for the example of Section 4.3, the composition of the optimal hedging program will be different from that using VaR as the risk measure (Table 5). Generally, the solution can vary greatly depending on the choice of risk measure (e.g., VaR, TVaR or standard deviation), the

parameter associated with the measures (e.g. percentile), or the mechanics of the underlying loss model (e.g., catastrophe model). For example, the optimal solution based on VaR would not necessarily be optimal if TVaR were used as the risk measure. Given the complex nature of this issue, using a coherent risk measure alone would not solve this problem. Although the parameter risk discussed above is not caused by or directly related to our optimization algorithm, the robustness of the outcome would be greatly improved if parameter risk could be handled more effectively, which remains a challenging problem for actuarial researchers and practitioners.

5. Summary

Index-based hedging instruments such as ILWs are increasingly recognized as effective hedging tools for insurance and reinsurance portfolios. However, wider adoption of these instruments is inhibited by basis risk, the random difference between the index-based payoff and the buyer's actual loss. This study presents a systematic approach for potential buyers to analyze and manage basis risk in order to take full advantage of the benefits offered by these instruments.

We examine two measures of basis risk: (i) hedging effectiveness and (ii) conditional payoff shortfall. Many existing measures such as the volatility of hedging (e.g., Major 1996) and R^2 (e.g., Harrington and Niehaus, 1999) are special cases of special cases of the hedging effectiveness measure, which quantifies the cost-adjusted benefit of the index-based hedging program relative to a benchmark. Next, we study the tradeoff between basis risk and the cost of hedging. The conditional payoff shortfall measures the probability that the buyer recovers less from an index-based hedging program than from a benchmark, reflecting the likelihood of "regret" for using the non-traditional hedging approach. In this study, a traditional catastrophe excess reinsurance layer is used as the benchmark. However, a wider spectrum of risk management products (such as proportional reinsurance, per risk excess reinsurance) is available. The methodology proposed in this paper is equally applicable to analyze these different benchmarks assuming the loss distribution of the underlying portfolio net of these products can be calculated.

Finally, we present a robust numerical algorithm designed to optimize an index-based hedging program consisting of multiple index-based contracts by analyzing the tradeoff between basis risk and cost of a hedging program. Compared to a benchmark optimization procedure based on exhaustive search, the GA-based approach is shown to work effectively to maximize the return on risk of a reinsurance portfolio subject to constraints. Nevertheless, like most financial optimization procedures, the outcome of this approach is not immune from parameter risk. Effectively address this issue remains a challenging but potentially rewarding future research direction.

Acknowledgement:

This paper would not have been possible without the encouragement, support and rigorous review by Bill Panning. The author would like to thank the CAS reviewers Tim Aman and David Drury for their extremely constructive input.

References:

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath, 1999, Coherent Measures of Risk, Journal of Mathematical Finance, 9(3), pp. 203-28.

Cummins, J. D., D. Lalonde, and R. D. Phillips, 2003: The basis risk of catastrophic-loss index securities, to appear in the *Journal of Financial Economics*.

Doherty, N.A. and A. Richter, 2002: Moral hazard, basis risk, and gap insurance. *The Journal of Risk and Insurance*, **69**(1), 9-24.

Goldberg, D.E., 1989: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Pub Co, 412pp.

Harrington S. and G. Niehaus, 1999: Basis risk with PCS catastrophe insurance derivative contracts. *Journal of Risk and Insurance*, **66**(1), 49-82.

Litzenberger, R.H., D.R. Beaglehole, and C.E. Reynolds, 1996: Assessing catastrophe reinsurance-linked securities as a new asset class. Journal of Portfolio Management, Special Issue Dec. 1996, 76-86.

Major, J.A., 1999: Index Hedge Performance: Insurer Market Penetration and Basis Risk, in Kenneth A. Froot, ed., *The Financing of Catastrophe Risk* (Chicago: University of Chicago Press).

Meyers, G.G., 1996: A buyer's guide for options and futures on a catastrophe index, Casualty Actuarial Society Discussion Paper Program, May, 273-296.

Zeng, L., 2000: On the basis risk of industry loss warranties, *The Journal of Risk Finance*, **1**(4) 27-32.

Martian Chronicles: Is MARS Better than Neural Networks?

Louise A. Francis, FCAS, MAAA

Martian Chronicles: Is MARS better than Neural Networks?

by Louise Francis, FCAS, MAAA

Abstract:
A recently developed data mining technique, Multivariate Adaptive Regression Splines (MARS) has been hailed by some as a viable competitor to neural networks that does not suffer from some of the limitations of neural networks. Like neural networks, it is effective when analyzing complex structures which are commonly found in data, such as nonlinearities and interactions. However, unlike neural networks, MARS is not a "black box", but produces models that are explainable to management.

This paper will introduce MARS by showing its similarity to an already well-understood statistical technique: linear regression. It will illustrate MARS by applying it to insurance fraud data and will compare its performance to that of neural networks.

Acknowledgements:
The author thanks the Automobile Insurers Bureau of Massachusetts for supplying the data used in this paper and wishes to acknowledge the efforts of Richard Derrig in providing insight into the data and approaches to analyzing it. The author also acknowledges the following people who reviewed this paper and provided many constructive suggestions: Jane Taylor, Francois Morrisette, Christopher Yaure, Patricia Francis-Lyon and Virginia Lambert.

Martian Chronicles: Is MARS better than Neural Networks?

The discipline of artificial intelligence has contributed a number of promising techniques to the analyst's toolkit. The techniques have names such as "machine learning", "genetic algorithms" and "neural networks". These techniques are collectively known as data mining. Data mining uses computationally intensive techniques to find patterns in data. When data mining tools are applied to data containing complex relationships they can identify relationships not otherwise apparent. These complexities have been a challenge for traditional analytical procedures such as linear regression.

The casualty actuarial literature contains only a few papers about data mining techniques. Speights *et al.* (Speights *et al.*, 1999) and Francis (Francis, 2001) introduced the neural network procedure for modeling complex insurance data. Hayward (Hayward, 2002) described the use of data mining techniques in safety promotion and better matching of premium rates to risk. The methods discussed by Hayward included exploratory data analysis using pivot tables and stepwise regression.

In this paper, a new technique, MARS, which has been proposed as an alternative to neural networks (Steinberg, 2001), will be introduced. The name MARS, coined for this technique by its developer, Freidman, (Hastie, *et al.*, 2001), is an acronym for Multivariate Adaptive Regression Splines. The technique is a regression based technique which allows the analyst to use automated procedures to fit models to large complex databases. Because the technique is regression based, its output is a linear function that is readily understood by analysts and can be used to explain the model to management. Thus, the technique does not suffer from the "black box" limitation of neural networks. However, the technique addresses many of the same data complexities addressed by neural networks.

Neural networks are one of the more popular data mining approaches. These methods are among of the oldest data mining methods and are included in most data mining software packages. Neural networks have been shown to be particularly effective in handling some complexities commonly found in data. Neural networks are well known for their ability to model nonlinear functions. The research has shown that a neural network with a sufficient number of parameters can model any continuous nonlinear function accurately.[1] Francis (Francis, 2001) also showed that neural networks are valuable in fitting models to data containing interactions. Neural networks are often the tools of choice when predictive accuracy is required. Berry and Linoff (Berry and Linoff, 1997) suggest that neural networks are popular because of their proven track record.

Neural networks are not ideal for all data sets. Warner and Misra presented several examples where they compared neural networks to regression (Warner and Misra, 1996). Their research showed that regression outperformed neural networks when the functional relationship between independent and dependent variables was known. Francis (Francis,

[1] A more technical description of the property is that with a sufficient number of nodes in the neural network's hidden layer, the neural network can approximate any deterministic nonlinear continuous function.

2001) showed that when the relationship between independent and dependent variables was linear, classical techniques such as regression and factor analysis outperformed neural networks.

Perhaps the greatest disadvantage of neural networks is the inability of users to understand or explain them. Because the neural network is a very complex function, there is no way to summarize the relationships between independent and dependent variables with functions that can be interpreted by data analysts or management. Berry and Linoff (Berry and Linoff, 1997) state that "Neural networks are best approached as black boxes with mysterious inner workings, as mysterious as the origins of our own consciousness". More conventional techniques such as linear regression result in simple mathematical functions where the relationship between predictor and target variables is clearly described and can be understood by audiences with modest mathematical expertise. The "black box" aspect of neural networks is a serious impediment to more widespread use.

Francis (Francis, 2001) listed several complexities found in actual insurance data and then showed how neural networks were effective in dealing with these complexities. This paper will introduce MARS and will compare and contrast how MARS and neural networks deal with several common data challenges. Three challenges that will be addressed in this paper are:

1) Nonlinearity: Traditional actuarial and statistical techniques often assume that the functional relationship between the independent variables and the dependent variable is linear or some transformation of the data exists that can be treated as linear.
2) Interactions: The exact form of the relationship between a dependent and independent variable may depend on the value of one or more other variables.
3) Missing data: Frequently data has not been recorded on many records of many of the variables that are of interest to the researcher.

The Data

This paper features the application of two data mining techniques, neural networks and MARS, to the fraud problem. The data for the application was supplied by the Automobile Insurers Bureau of Massachusetts (AIB). The data consists of a random sample of 1400 closed claims that were collected from PIP (personal injury protection or no-fault coverage) claimants in Massachusetts in 1993. The database was assembled with the cooperation of ten large insurers. This data has been used by the AIB, the Insurance Fraud Bureau of Massachusetts (IFB) and other researchers to investigate fraudulent claims or probable fraudulent claims (Derrig *et al.*, 1994, Weisberg and Derrig, 1995, Viaene *et al.*, 2002). While the typical data mining application would use a much larger database, the AIB PIP data is well suited to illustrating the use of data mining techniques in insurance. Viaene *et al.* used the AIB data to compare the performance of a number of data mining and conventional classification techniques (Viaene *et al.*, 2002).

Two key fraud related dependent variables were collected in the study: an overall assessment (ASSESS) of the likelihood the claim is fraudulent or abusive and a suspicion score (SUSPICION). Each record in the data was assigned a value by an expert. The value indicates the expert's subjective assessment as to whether the claim was legitimate or whether fraud or abuse was suspected. Experts were asked to classify suspected fraud or abuse claims into the following categories: exaggerated damages, opportunistic fraud or planned fraud. As shown in Table 1, the assessment variable can take on 5 possible values. In addition, each claim was assigned a score from 0 (none) to 10 (very high) indicating the expert's degree of suspicion that the claim was abusive or fraudulent. Weisberg and Derrig (Weisberg and Derrig, 1993) found that more serious kinds of fraud, such as planned fraud were associated with higher suspicion scores than "softer" fraud such as exaggeration of damages. They suggest that the suspicion score was able to measure the range of "soft" versus "hard" fraud.

The database contains detailed objective claim information on each claim in the study. This includes information about the policy inception date, the date the accident occurred, the date it was reported, the paid and incurred loss dollars, the injury type, payments to health care providers and the provider type. The database also contains "red flag" or fraud indicator variables. These variables are subjective assessments of characteristics of the claim that are believed to be related to the likelihood of fraud or abuse. More information on the variables in the model is supplied below in the discussion of specific models.

Table 1

Assessment Variable		
Value	**Assessment**	**Percent of Data**
1	Probably legitimate	64%
2	Excessive treatment only	20%
3	Suspected opportunistic fraud, no injury	3%
4	Suspected opportunistic fraud, exaggerated injury	12%
5	Suspected planned fraud	1%

We may use the more inclusive term "abuse" when referring to the softer kinds of fraudulent activity, as only a very small percentage of claims meet the strict standard of criminal fraud (Derrig, 2002). However, misrepresentation and exaggeration of the nature and extent of the damages, including padding of the medical bills so that the value of the claim exceeds the tort threshold, occur relatively frequently. While these activities are often thought of as fraud, they do not meet a legal definition of fraud. Therefore, they will be referred to as abuse. Overall, about one third of the claims were coded as probable abuse or fraud claims.

Nonlinear Functions

The relationships encountered in insurance data are often nonlinear. Classical statistical modeling methods such as linear regression have had a tremendous impact on the analysis and modeling of data. However, traditional statistical procedures often assume

that the relationships between dependent and independent variables are linear. Traditional modeling also allows linear relationship that result from a transformation of dependent or independent variables, so some nonlinear relationships can be approximated. In addition, there are techniques specifically developed for fitting nonlinear functions such as nonlinear regression. However, these techniques require that theory or experience specify the "true" form of the nonlinear relationships. Data mining techniques such as neural networks and MARS do not require that the relationships between predictor and dependent variables be linear (whether or not the variables are transformed). Both neural networks and MARS are also considered nonparametric because they require no assumptions about the form of the relationship between dependent and independent variables.

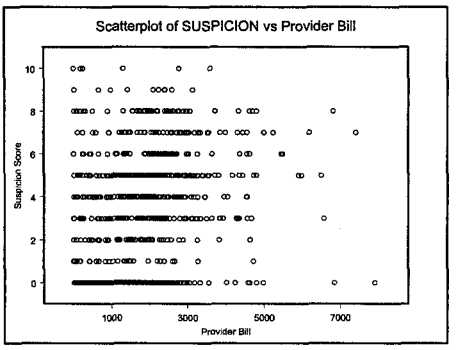
For this illustration, a dependent variable that is not categorical (i.e. values have a meaningful order) was selected. The selected dependent variable was SUSPICION. Unlike the ASSESS variable, the values on the SUSPICION variable have a meaningful range, with higher values associated with suspicion of more serious fraud.

To illustrate methods of fitting models to nonlinear curves, a variable was selected which 1) had a significant correlation with the dependent variable, and 2) displayed a highly nonlinear relationship. Illustrating the techniques is the objective of this example. The data used may require significant time to collect and may therefore not be practical for an application where the objective is to predict abuse and fraud (which would require data that is available soon after the claim is reported). Later in the paper, models for prospectively predicting fraud will be presented. The variable selected was the first medical provider's bill[2]. A medical provider may be a doctor, a clinic, a chiropractor or a physical therapist. Prior published research has indicated that abusive medical treatment patterns are often key drivers of fraud (Derrig *et al.*, 1994, Weisberg and Derrig, 1995). Under no-fault laws, claimants will often deliberately run the medical bills up high enough to exceed tort thresholds. In this example the relationship between the first provider's medical bill and the value of the suspicion score will be investigated. The AIB fraud database contains the medical bills submitted from the top two health care providers. If more costly medicine is delivered to suspicious claims than non-suspicious claims, the provider bills should be higher for the suspicious claims.

Figure 1 presents a scatterplot of the relationship between SUSPICION and the provider bill. No relationship is evident from the graph. However, certain nonlinear relationships can be difficult to detect visually.

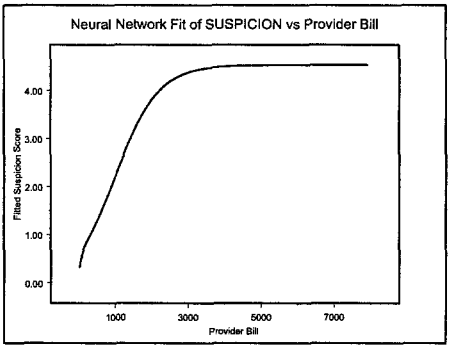
[2] Note that Massachusetts PIP covers only the first $8,000 of medical payments if the claimant has health insurance. Large bill amounts may represent data from claimants with no coverage. Bills may also exceed $8,000 even if payments are limited. However, the value of medical bills on some claims may be truncated because reimbursement is not expected.

Figure 1



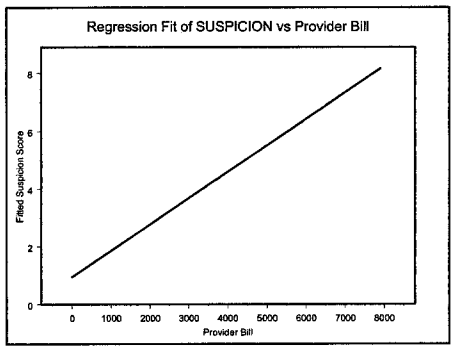
Neural networks will first be used to fit a curve to the data. A detailed description of how neural networks analyze data is beyond the scope of this paper. Several sources on this topic are Francis, Lawrence and Smith (Francis, 2001, Lawrence, 1994, Smith, 1996). Although based upon how neurons function in the brain, the neural network technique essentially fits a complex non-parametric nonlinear regression. A task at which neural networks are particularly effective is fitting nonlinear functions. The graph below displays the resulting function when the dependent variable SUSPICION is fit to the provider bill by a neural network. This graph displays a function that increases quickly at lower bill amounts and then levels off. Although the curve is flat over much of the range of medical bills, it should be noted that the majority of bills are below $2,000 (in 1993 dollars).

Figure 2



One of the most common statistical procedures for curve fitting is linear regression. Linear regression assumes the relationship between the dependent and independent variables is linear. Figure 3 displays the graph of a fitted regression line of SUSPICION on provider bill. The regression forces a linear fit to SUSPICION versus the payment amount. Thus, rather than a curve with a rapidly increasing trend line that levels off, a line with a constant slope is fitted. If the relationship is in fact nonlinear, this procedure is not as accurate as that of the neural network.

Figure 3

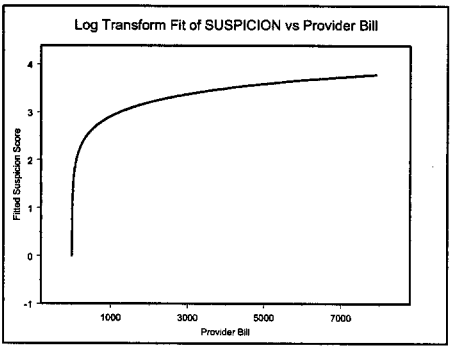


When the true relationship between a dependent and independent variable is nonlinear, various approaches are available when using traditional statistical procedures for fitting the curve. One approach is to apply a nonlinear transformation to the dependent or independent variable. A linear regression is then fit to the transformed variables. As an example, a log transform was applied to the provider bill variable in the AIB data. The regression fit was of the form:

$$Y = B_o + B_1 \ln(X)$$

That is, the dependent variable, the suspicion score, is assumed to be a linear function of the natural log of the independent variable, provider bill. Figure 4 displays the curve fit using the logarithmic transformation.

Figure 4

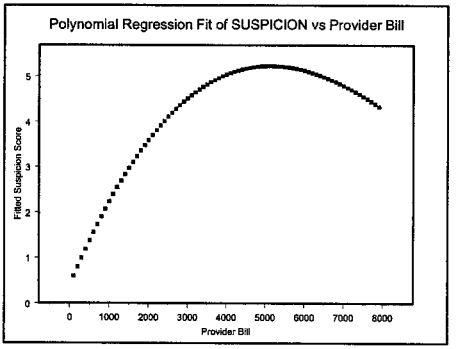


Another procedure which is used in classical linear regression to approximate nonlinear curves is polynomial regression. The curve is approximated by the function:

$$Y = B_o + B_1 X + B_2 X^2 + ... + B_n X^n$$

Generally, low order polynomials are used in the approximation. A cubic polynomial (including terms up to provider bill raised to the third power) was used in the fit. Figure 5 displays a graph of a fitted polynomial regression.

Figure 5



The use of polynomial regression to approximate functions is familiar to readers from its use in Taylor series expansions for this purpose. However, the Taylor series expansion is used to approximate a function near a point, rather than over a wide range. When evaluating a function over a range, the maximums and inflection points of the polynomial may not exactly match the curves of the function being approximated.

The neural network model had an R^2 (coefficient of determination) of 0.37 versus 0.25 for the linear model and 0.26 for the log transform. The R^2 of the polynomial model was comparable to that of the neural network model. However, the fit was influenced strongly by a small number of claims with large values. Though not shown in the graph, at high values for the independent variable the curve declines below zero and then increases again. This unusual behavior suggests that the fitted curve may not approximate the "true" relationship between provider bill and suspicion score well at the extremes of the data and may perform poorly on new claims with values outside the range of the data used for fitting.

Table 2 below shows the values of SUSPICION for ranges of the provider bill variable. The table indicates that SUSPICION increases rapidly at low bill amounts and then levels off at about $3,000.

Table 2

Suspicion Scores by Provider Bill

Provider Bill	Number of Claims	Mean Suspicion Score
$0	444	0.3
1 - 1,000	376	1.1
1,001 - 2,000	243	3.0
2,001 - 3,000	227	4.2
3,001 - 4,000	60	4.6
4,001 - 5000	33	4.2
5,001 - 6,000	5	5.8
6,001 - 7,000	12	4.3

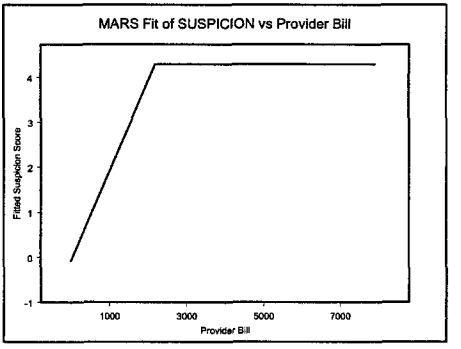
The examples illustrate that traditional techniques which require specific parametric assumptions about the relationship between dependent and independent variables may lack the flexibility to model nonlinear relationships. It should be noted, however, that Francis (Francis, 2001) presented examples where traditional techniques performed as well as neural networks in fitting nonlinear functions. Also, when the true relationship between the dependent and independent variables is linear, classical statistical methods are likely to outperform neural networks.

MARS and Nonlinear Functions

The MARS approach to fitting nonlinear functions has similarities to polynomial regression. In its simplest form MARS fits piecewise linear regressions to the data. That is, MARS breaks the data into ranges and allows the slope of the line to be different for the different ranges. MARS requires the function fit to be continuous, thus there are no jump points between contiguous ranges.

To continue the previous example, a function was fit by MARS. The graph below displays the MARS fitted function. It can be seen that the curve is broken into a steeply sloping line, which then levels off much the way the neural network fitted function did.

Figure 6



MARS uses an optimization procedure that fits the best piecewise regression. Simpler functions may adequately approximate the relationship between predictor and dependent variables and are favored over more complex functions. From the graph, it can be seen that the best MARS regression had two pieces:

1) The curve has a steep slope between bill amounts of $0 and $2,185
2) The curve levels off at bill amounts above $2,185

The fitted regression model can be written as follows:

$BF1 = \max(0, 2185 - X)$
$Y = 4.29 - 0.002 * BF1$

where

Y is the dependent variable (Suspicion score)
X is the provider bill

The points in the data range where the curves change slope are known as knots. The impact of knots on the model is captured by basis functions. For instance BF1 is a basis function. Basis functions can be viewed as similar to dummy variables in linear regression. Dummy variables are generally used in regression analysis when the predictor variables are categorical. For instance, the Provider bill variable can be

converted into a categorical variable by using amount ranges for the categories. We could have the following categories:

Range 1	$0 - $2,185	Dummy Variable = 1
Range 2	> $2,185	Dummy Variable = 0

A dummy variable is a binary indicator variable. It will have a value of 1 when the bill falls within the specified interval for the dummy. Here if the bill is $1,000, D1 will be 1. When it is $5,000 D1 will be 0.

A regression with dummy variables has the form:

$$Y = B_0 + B_1*D1 + B_2 * D2 + B_3*D3+ \ldots + B_n * Dn$$

Since in this simple example there is only one dummy variable, the model is:

$$Y = B_0 + B_1*D1$$

The constant B_0 captures the effect of the first or base group (greater than $2185). The dummy variable D1 captures the effect of its bill group relative to the base group. The coefficients for the above model when fitted to the AIB data were:

$$Y = 4.28 - 2.89*D1$$

This regression function indicates that the mean suspicion score is 4.28 for bills greater than $2,185 and 1.39 for bills less than $2,185. However, the use of categorical dummy variables (as opposed to basis functions) creates jumps in the level of the dependent variable, rather than a linear curve, when the range changes.

Basis Functions and Dummy Variables
Each basis function is a combination of a dummy variable with a continuous variable. In the regression function between suspicion score and provider bill:

$$BF1 = \max(0, 2185 - X)$$

$$Y = 4.287 - 0.002 * BF1$$

BF1 can be rewritten as:

$$BF1 = D1*(2185 - X)$$

where D1 is a dummy variable, which takes on the value of 0 if the provider bill is greater than or equal to $2,185 and 1 if it is less than that value.

Finding the Knots

As mentioned above, a knot is the point in a range at which the slope of the curve changes. Both the number of knots and their placement are unknown at the beginning of the process. A stepwise procedure is used to find the best points to place the spline knots. In its most general form each value of the independent variable is tested as a possible point for placement of a knot. The model initially developed is overfit. A statistical criterion that tests for a significant impact on a goodness of fit measure is used to remove knots. Only those that have a significant impact on the regression are retained. The statistical criterion, generalized cross-validation, will be described later in the paper.

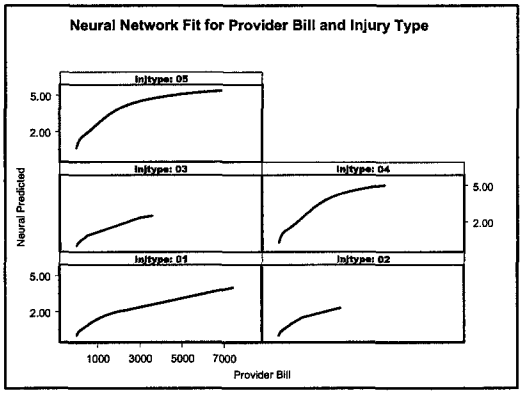
Fitting Smooth Curves

The above discussion describes spline functions which are piecewise linear regressions. For such regressions there is a break in the slope of the curve at each knot. A smooth curve could be created by allowing for higher order terms in the regression, i.e. quadratic or cubic terms could be included. Often, when fitting smoothing splines the curve is a cubic curve. For cubic splines, there is a requirement that the first and second derivatives are continuous at the knot points. For the remainder of this paper we will use piecewise linear splines. Although cubic splines produce smoother curves, they do not, in general, (Steinberg, 1999) significantly improve the fit of the model and are more difficult to parameterize.

Functions with Interaction Terms

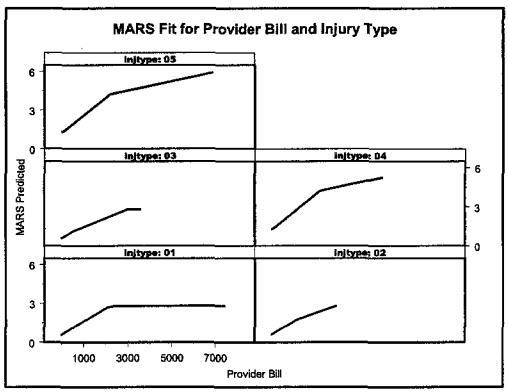
The illustrations shown so far demonstrate MARS's capability for modeling nonlinear relationships. Another complication that occurs when working with insurance data is known as "interactions". The relationship between a predictor variable and the target variable may depend on the value of a second variable. For instance, the relationship between the medical provider bill and the suspicion score may depend on the injury type. This hypothesis is supported by the results of fitting a neural network model for SUSPICION to provider bill and injury type (shown in Figure 7). (For presentation purposes, only some of the injury types are shown). It can be seen that the curves for injury type 4 (neck sprain), and type 5 (back sprain) increase faster than those of the other injury types and ultimately plateau at higher levels.

Figure 7



A MARS curve was fit to the fraud interaction data. The results of the fit are shown below:

Figure 8



It can be seen that, as with the neural network, injury type 4 (neck sprain), and type 5 (back sprain) increase faster and have higher scores than the other injury types. The MARS fitted function was:

BF1 = max(0, 2185 - X)
BF2 = (INJTYPE = 4 OR INJTYPE = 5)
BF3 = max(0, X - 159) * BF2
Y = 2.815 - 0.001 * BF1 + 0.685 * BF2 + .360E-03 * BF3

where
X is the provider bill
INJTYPE is the injury type

There are three basis functions in the model. Basis function BF1 splits the provider bill into the range $0 to $2,185 and greater than $2,185. It is like the first basis function in the previous model of SUSPICION and provider bill. Basis function BF2 is a categorical dummy variable, based on the value of injury type. If the injury type is 4 or 5, it takes on a value of 1, otherwise it is 0. In the model, the coefficient of BF2 is 0.685. Thus, the regression constant value is increased by 0.685 if the injury is a sprain. Basis function BF3 captures the interaction between injury type and provider bill and increases the slope of the curve for sprains.

To create the BF2 basis function, MARS searches all the categories of injury type. By recursive partitioning, or sequential splitting of the categories into two distinct groups, it groups together those categories with a similar effect on the dependent variable into basis functions. When there is more than one categorical variable, the procedure is performed on each one. Only those basis functions with a significant effect on the target variable, as determined by the improvement in the R^2, are included in the final model.

Similarly, an automated search procedure is used to create basis functions that specify interaction effects. Combinations of predictors are tested two at a time for two-way interaction[3]. New basis functions may be created to capture the interaction effect. Thus, a different combination of the injury types than those in BF2 could be associated with the interaction of injury type and provider bill. For this model the injury types were the same for BF2 and BF3.

This example illustrates one advantage of MARS over other data mining techniques such as neural networks. MARS groups together related categories of nominal variables. Many insurance categorical variables have many different levels[4]. For instance, while the injury type variable in the AIB data has only 15 levels, injury type data often has hundreds or even thousands of possible values. Increasingly, the insurance industry is shifting to the use of ICD9[5] codes for injuries. There are in excess of 15,000 possible

[3] Higher order interactions, such as three way and four way interactions are permissible. However, high order interactions are unlikely to be statistically significant in a database of this size.
[4] Note that another data mining technique, Decision Trees (also know as CART) can also group together categories with similar impacts on the dependent variable.
[5] ICD9 codes are the codes used by medical providers and health insurers to classify injuries and illnesses. The definition of the classes is standardized and there is widespread use of these codes.

values for ICD9, many of which are related to similar illnesses or injuries. A procedure that can group together codes with a similar impact on the dependent variable is very handy when so many values are available. The neural network procedure turns each of the possible values of a categorical variable into a binary dummy variable when fitting a model. Many of these categories contain a tiny fraction of the data, thus the parameters fitted to the categories of the categorical variables may be very unstable. Collapsing the categories into a smaller number, with each group having a similar impact on the dependent variable (perhaps when interacting with another variable) significantly reduces the number of parameters in the model.

Missing Data

Missing data occurs frequently when working with large databases. The software commonly used for applying statistical models (including neural networks) typically applies very crude rules when data is missing. Such rules include elimination of records where any value on any variable is missing and substitution of the mean of a variable for the missing value on that variable. More sophisticated methods for addressing missing values, such as data imputation and the expectation maximization (EM) algorithm, have been developed. However, these methods are not widely available in the major statistical software packages. Two significant problems occur with missing data.

1. Because many statistical packages eliminate any record with a missing value on any variable, a lot of the data can be lost to the analysis.
2. In order for the analysis to be valid, the analyst must assume that value of both the dependent and predictor variables is independent of the presence of missing values.

MARS handles missing data by creating a basis function for any variable with missing data. This variable has a value of one when the data is missing on a given variable and zero otherwise. The search procedure can then determine if an interaction between missing data basis functions and other variables in the data are significant in predicting the dependent variable. Thus, other variables can act as surrogates for the missing variable.

Neural networks were not developed with the treatment of missing data in mind. Many neural network software products automatically eliminate from the model any record with a missing value for any variable in the model. Nevertheless there are procedures that can be used to deal with this challenge. One approach is to assign a constant value to data missing on a variable. This value is often the mean for that variable, but this need not be the case. Because neural networks fit nonlinear functions, a value not in the range of the remainder of the data might be assigned to the missing data on a variable, allowing a different relationship between independent and dependent variable for this value than for the remainder of the data. In addition, a dummy variable can be constructed for each of the variables with missing data, and this can be used in the neural network model. Unfortunately, the software available for fitting neural networks does not provide an automated approach to addressing the missing data issue so significant additional programming effort may be required.

To illustrate the handling of missing data, suspicion score is modeled as a function of total provider medical bill and health insurance. The total provider medical bill is the sum of the bills from all providers. Health insurance is a categorical variable with values of yes (claimant has health insurance), no (claimant does not have health insurance) and unknown (missing). The table below shows the distribution of each of these values in the data. The variables in this example were selected because they provided a good illustration of the handling of missing values. That is, the health insurance variable had a significant number of missing cases (see table below) and the total medical bill's influence on the dependent variable is impacted by the presence/absence of missing values on this variable.

Table 3
Health Insurance

Value	Frequency	Percent	Cumulative Percent
No	457	32.6	32.6
Missing	208	14.9	47.5
Yes	735	52.5	100.0
Total	1400	100.0	

The following MARS model was fit:

BF1 = max(0, MP_BILL - 2885)
BF2 = max(0, 2885 - MP_BILL)
BF3 = (HEALTHIN ≠ MISSING)
BF4 = (HEALTHIN = MISSING)
BF5 = (HEALTHIN = N)
BF7 = max(0, MP_BILL - 2262) * BF5
BF8 = max(0, 2262 - MP_BILL) * BF5
BF9 = max(0, MP_BILL - 98) * BF4
BF10 = max(0, 98 - MP_BILL) * BF4
BF11 = max(0, MP_BILL - 710) * BF3
BF13 = max(0, MP_BILL - 35483)
BF15 = BF3 * BF2

Y = -0.754 - 0.002 * BF1 + 0.967 * BF3 + 1.389 * BF5 - .808E-04 * BF7
- .624E-03 * BF8 + 0.001 * BF9 + 0.016 * BF10
+ 0.001 * BF11 + .114E-03 * BF13 + .376E-03 * BF15

where:
MP_BILL is the total provider medical bill
HEALTHIN is the health insurance variable
BF1 – BF15 are the basis functions
Y is the dependent variable, suspicion score

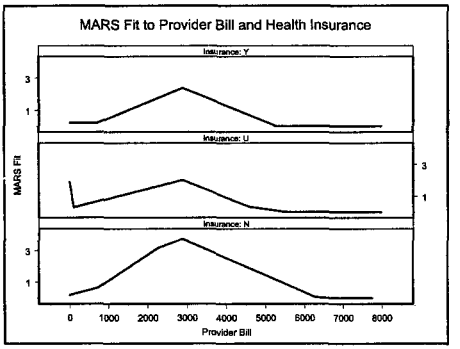
Note that there is no BF6, BF12 or BF14. Variables BF6, BF12 and BF14 were created by MARS, but as they were not found to be significant, they were not included in the final model.

The MARS model created two basis functions for the missing values, one for the presence of missing values and one for the absence. It can be seen that the shape and level of the curve depends on both the value of the health insurance variable and whether it is missing. Basis functions BF3 and BF4 are the dummy variables denoting missing/not-missing values on the health insurance variable. If the health insurance information is missing, BF4 is one. If the information is not missing, BF3 is one. The model indicates that the overall score is raised by .967 if health insurance information is present. Basis functions BF9 and BF10 are the interactions of a missing value on health insurance with provider bill. Basis functions BF11 and BF15 are the interaction of health insurance not missing with total provider bill. Thus, when the provider bill is less than $98 and the health insurance information is missing, the curve's slope is increased by 0.016. This causes the suspicion score to spike at low provider bill values. BF11 indicates that the slope of the curve increases by .001 for values above $710 and BF15 indicates that the slope of the curve increases by 0.00038 up to bill values of $2,885, when health insurance information is present.

Figure 9 displays the curve fit by MARS.[6] The top graph is curve for health insurance (i.e. equal to "yes"), the middle curve is the curve for health insurance unknown (missing) and the bottom graph is the curve for no health insurance. The figure shows that suspicion scores are on average highest when the claimant does not have health insurance and lowest when the information about health insurance is missing. The graphs show that suspicion scores for all categories decline after values of about $3,000.

[6] In the graph, suspicion scores of less than one were censored to have a value of zero.

Figure 9



A neural network was fit to the data using the dummy variable approach described above. That is, a dummy variable was created for the presence or absence of a value on the health insurance variable. Figure 10 shows a comparison of the MARS and the neural network fitted values. The curves fit by the neural network did not vary much over the different values of the health insurance variable. Moreover, for health insurance missing and health insurance equal to 'Y' the neural network scores are above the MARS scores for provider bills greater than about $1,000. In addition, the MARS model suspicion scores decline at high bill amounts, but they do not for the neural network model. Table 4 presents average suspicion scores by bill amount categories for each of the values on the health insurance variable. This table indicates that suspicion scores are higher for claimants with health insurance information, and are highest for claimants with no health insurance. The table also indicates that the suspicion score declines at higher bill amounts, but the decline in the data seems to occur later than the MARS model indicates.

Figure 10

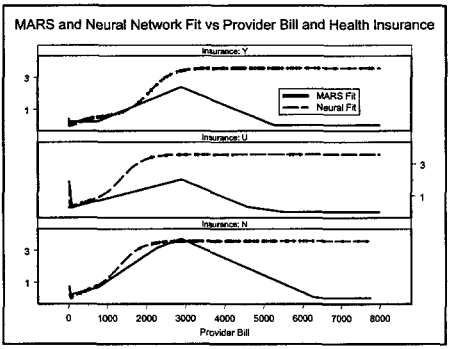


Table 4

		Suspicion Scores by Health Insurance Category		
Total Provider Bill	**Claim Count**	**Y**	**U**	**N**
$0	65	0.3	2.1	0.7
1 - 1,000	532	0.3	0.5	0.4
1,001 - 2,000	140	1.2	3.1	1.9
2,001 - 3,000	268	2.9	3.0	4.5
3,001 - 4,000	149	3.1	2.9	4.2
4,001 - 5000	85	3.4	-	4.8
5,001 - 6,000	54	3.0	2.5	3.4
6,001 - 7,000	25	4.4	-	5.1
7.001 - 8,000	18	2.6	-	4.5
8,001 - 9,000	12	2.8	4.0	-
9,001 - 11,000	13	3.1	-	2.7
> 11,000	39	1.0	-	2.5
Total	1,400	1.6	1.5	2.8

Both the MARS model and the neural network model had similar R^2 (around 0.37). The MARS software uses a statistical procedure to assess the significance of variables and rank them in order of importance. This procedure is described in a later section of this paper. The MARS procedure found the health insurance variable to be significant, but much less significant than the provider bill variable. By visual inspection, it appears that the neural network procedure found no meaningful difference in suspicion score by health insurance category. A more formal neural network procedure for assessing the importance of variables will be discussed in the next section of the paper.

A simple procedure for comparing the accuracy of two models is to hold out a portion of the data for testing. Data is separated into training and test data. The model is fit using the training data and its accuracy is tested using the test data to determine how well the dependent variable was predicted on data not used for fitting. This test is relatively straightforward to perform. In the next section of the paper a more computationally intensive procedure will be presented.

To compare the neural network and MARS models, two thirds of the data was used for fitting and one third was used for testing. The neural network had an R^2 of 0.30 compared to 0.33 for the MARS model. The performance of the two models was also tested on subsets of the data containing only one value of the health insurance variable (i.e., health insurance missing, health insurance equal to yes and health insurance equal to no). MARS outperformed the neural network model on health insurance missing (R^2 = .26 versus $R^2 = 0$) and health insurance equal to no (R^2 = .31 versus R^2= .25). The neural network outperformed MARS on health insurance equal to yes (R^2= .43 versus R^2 = .32).

This example suggests that MARS more accurately modeled the effect of the health insurance variable and the effect of a missing value for this variable on the dependent variable than did the neural network model. However, it would be desirable to assess the significance of the differences in the accuracy of the overall fit.

The square root of R^2 is the correlation coefficient, which can be used in a test of significance. The distribution of a transform of the correlation coefficient can be approximated by a normal distribution[7]:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

$$\mu_z = Z,$$

$$\sigma_z = \frac{1}{n-3}$$

where r is the correlation coefficient and n is the sample size.

The normal approximation was used to compute confidence intervals for each of the correlations. As shown in Table 5, the 95% confidence intervals around the Z statistic computed from the two correlations overlapped, suggesting that the difference between the fits of the two models is not statistically significant.

Table 5

Confidence Intervals for Correlation Coefficient

Model	R^2	r	Z	sd	Lower 95% CI	Upper 95% CI
MARS	0.33	0.57	0.65	0.05	0.56	0.74
Neural Network	0.30	0.55	0.62	0.05	0.52	0.71

This example illustrates one of the great strengths of MARS: its automated procedures for handling missing data. While missing data was not a major issue with the AIB database, as most of the variables were fully populated, it is a common problem with most insurance databases. One possible use for MARS is to create basis functions for variables having missing values. These basis functions could then be used by other procedures such as neural networks.

A More Complex Model

The models presented thus far have been relatively simple one and two variable models. In this section of the paper, the results of a more complex model will be presented. The variables used in the model are described below.

This section will present an example where MARS and neural networks are used for classification. The dependent variable for this model is ASSESS, the expert's assessment of the likelihood that the claim is a fraud or abuse claim. This variable was converted to a binary dependent variable. The two categories were the value 1 (probably legitimate) versus 2 through 5 (the various kinds of suspected fraud or abuse). Thus, if a claim is other than probably legitimate, it is treated as a suspected abuse claim.

[7] This formula is from Miller and Wichern (Miller and Wichern, 1977, pp. 213 – 214).

MARS can perform regressions on binary variables. When the dependent variable is binary, MARS is run in binary mode. In binary mode, the dependent variable is converted into a 0 (legitimate) or a 1 (suspected fraud or abuse). Ordinary least squares regression is then performed regressing the binary variable on the predictor variables. Logistic regression is a more common procedure when the dependent variable is binary. Suppose that the true target variable is the probability that a given claim is abusive, and this probability is denoted $p(x)$. The model relating $p(x)$ to the a vector of independent variables **x** is:

$$\ln(\frac{p}{1-p};\mathbf{x}) = B_0 + B_1X_1 + ... + B_nX_n$$

where the quantity $\ln(p(\mathbf{x})/(1-p(\mathbf{x})))$ is known as the logit function or log odds. Logistic regression can be used to produce scores that are between zero and one, consistent with viewing the score as a probability. Binary regressions can produce predicted values which can be less than zero and greater than one. One solution to this issue is to truncate the predicted values at zero and one. Another solution is to add the extra step of fitting a logistic regression to the data using the MARS predicted value as the independent variable and the binary assessment variable as the dependent variable. The fitted probabilities from the logistic regression can then be assigned as a score for the claim. The neural network model was also run in binary mode and also produced fitted values which were less than zero or greater than one. In this analysis, logistic regression was applied to the results of both the MARS and neural network fits to convert the predicted values into probabilities.

Variables in the Model

There are two categories of predictor variables that were incorporated into the models described in this section. The first category is red flag variables. These are primarily subjective variables that are intended to capture features of the accident, injury or claimant that are believed to be predictive of fraud or abuse. Many red flag variables represent accumulated industry wisdom about which indicators are likely to be associated with fraud or abuse. The information recorded in these variables represents an expert's subjective assessment of fraud indications, such as "the insured felt set up, denied fault". These variables are binary, that is, they are either true or false. Such red flag variables are often used to target certain claims for further investigation. The data for these red flag variables is not part of the claim file; it was collected as part of the special effort undertaken in assembling the AIB database for fraud research.

The red flag variables were supplemented with claim file variables deemed to be available early in the life of a claim and therefore of practical value in predicting fraud and abuse.

The variables selected for use in the full model are the same as those used by Viaene *et al.* (Viaene *et. al.*, 2002) in their comparison of statistical and data mining methods. While a much larger number of predictor variables is available in the AIB data for

modeling fraud, the red flag and objective claim variables selected for incorporation into their models by Viaene *et al.* were chosen because of early availability. Therefore they are likely to be useful in predicting fraud and abuse soon enough in the claim's lifespan for effective mitigation efforts to lower the cost of the claim. Tables 6 and 7 present the red flag and claim file variables.

Table 6

Red Flag Variables

Subject	Indicator Variable	Description
Accident	ACC01	No report by police officer at scene
	ACC04	Single vehicle accident
	ACC09	No plausible explanation for accident
	ACC10	Claimant in old, low valued vehicle
	ACC11	Rental vehicle involved in accident
	ACC14	Property Damage was inconsistent with accident
	ACC15	Very minor impact collision
	ACC16	Claimant vehicle stopped short
	ACC19	Insured felt set up, denied fault
Claimant	CLT02	Had a history of previous claims
	CLT04	Was an out of state accident
	CLT07	Was one of three or more claimants in vehicle
Injury	INJ01	Injury consisted of strain or sprain only
	INJ02	No objective evidence of injury
	INJ03	Police report showed no injury or pain
	INJ05	No emergency treatment was given
	INJ06	Non-emergency treatment was delayed
	INJ11	Unusual injury for auto accident
Insured	INS01	Had history of previous claims
	INS03	Readily accepted fault for accident
	INS06	Was difficult to contact/uncooperative
	INS07	Accident occurred soon after effective date
Lost Wages	LW01	Claimant worked for self or a family member
	LW03	Claimant recently started employment

Table 7

Claim Variables Available Early in Life of Claim

Variable	Description
AGE	Age of claimant
POLLAG	Lag from policy inception to date of accident[8]
RPTLAG	Lag from date of accident to date reported
TREATLAG	Lag from date of accident to earliest treatment by service provider
AMBUL	Ambulance charges
PARTDIS	The claimant partially disabled
TOTDIS	The claimant totally disabled
LEGALREP	The claimant represented by an attorney

[8] POLLAG, RPTLAG and TRTLAG are continuous variables.

One of the objectives of this research is to investigate which variables are likely to be of value in predicting fraud and abuse. To do this, procedures are needed for evaluating the importance of variables in predicting the target variable. Below, we present some methods that can be used to evaluate the importance of the variables.

Evaluating Variable Importance

A procedure that can be used to evaluate the quality of the fit when fitting complex models is generalized cross-validation (GCV). This procedure can be used to determine which variables to keep in the model, as they produce the best fit, and which to eliminate. Generalized cross-validation can be viewed as an approximation to cross-validation, a more computationally intensive goodness of fit test described later in this paper.

$$GCV = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \hat{f}(x_i)}{1 - k/N}\right]^2$$

where N is the number of observations
y is the dependent variable
x is the independent variable(s)
k is the effective number of parameters or degrees of freedom in the model.

The effective degrees of freedom is the means by which the GCV error functions puts a penalty on adding variables to the model. The effective degrees of freedom is chosen by the modeler. Since MARS tests many possible variables and possible basis functions, the effective degrees of freedom used in parameterizing the model is much higher than the actual number of basis function in the final model. Steinberg states that research indicates that k should be two to five times the number of basis functions in the model, although some research suggests it should be even higher (Steinberg, 2000).

The GCV can be used to rank the variables in importance. To rank the variables in importance, the GCV is computed with and without each variable in the model.

For neural networks, a statistic known as the sensitivity can be used to assess the relative importance of variables. The sensitivity is a measure of how much the predicted value's error increases when the variables are excluded from the model one at a time. Potts (Potts, 2000) and Francis (Francis, 2001) described a procedure for computing this statistic. Many of the major data mining packages used for fitting neural networks supply this statistic or a ranking of variables based on the statistic. Statistical procedures for testing the significance of variables are not well developed for neural networks. One approach is to drop the least important variables from the model, one at a time and evaluate whether the fit deteriorates on a sample of claims that have been held out for testing. On a large database this approach can be time consuming and inefficient, but it is feasible on small databases such as the AIB database.

Table 8 displays the ranking of variable importance from the MARS model. Table 9 displays the ranking of importance from the neural network model. The final model fitted by MARS uses only the top 12 variables in importance. These were the variables that were determined to have made a significant contribution to the final model. Only variables included in the model, i.e., found to be significant are included in the tables.

Table 8

MARS Ranking of Variables

Rank	Variable	Description
1	LEGALREP	Legal Representation
2	TRTMIS	Treatment lag missing
3	ACC04	Single vehicle accident
4	INJ01	Injury consisted of strain or sprain only
5	AGE	Claimant age
6	PARTDIS	Claimant partially disabled
7	ACC14	Property damage was inconsistent with accident
8	CLT02	Had a history of previous claims
9	POLLAG	Policy lag
10	RPTLAG	Report lag
11	AMBUL	Ambulance charges
12	ACC15	Very minor impact collision

The ranking of variables as determined by applying the sensitivity test to the neural network model is shown below.

Table 9

Neural Network Ranking of Variables

Rank	Variable	Description
1	LEGALREP	Legal Representation
2	TRTMIS	Treatment lag missing
3	AMBUL	Ambulance charges
4	AGE	Claimant age
5	PARTDIS	Claimant partially disabled
6	RPTLAG	Report lag
7	ACC04	Single vehicle accident
8	POLLAG	Policy lag
9	CLT02	Had a history of previous claims
10	INJ01	Injury consisted of strain or sprain only
11	ACC01	No report by police officer at scene
12	ACC14	Property damage was inconsistent with accident

Both the MARS and the neural network find the involvement of a lawyer to be the most important variable in predicting fraud and abuse. Both procedures also rank as second a missing value on treatment lag. The value on this variable is missing when the claimant has not been to an outpatient health care provider, although in over 95% of these cases,

the claimant has visited an emergency room.[9] Note that both medical paid and total paid for this group is less than one third of the medical paid and total paid for claimants who visited a provider. Thus the TRTMIS (treatment lag missing) variable appears to be a surrogate for not using an outpatient provider. The actual lag in obtaining treatment is not an important variable in either the MARS or neural network models.

Explaining the Model

Below are the formulas for the model fit by MARS. Again note that some basis functions created by MARS were found not to be significant and are not shown. To assist with interpretation, Table 10 displays a description of the values of some of the variables in the model.

```
BF1 = (LEGALREP = 1)
BF2 = (LEGALREP = 2)
BF3 = ( TRTLAG = missing)
BF4 = ( TRTLAG ≠ missing)
BF5 = ( INJ01 = 1) * BF2
BF7 = ( ACC04 = 1) * BF4
BF9 = ( ACC14 = 1)
BF11 = ( PARTDIS = 1) * BF4
BF15 = max(0, AGE - 36) * BF4
BF16 = max(0, 36 - AGE ) * BF4
BF18 = max(0, 55 - AMBUL ) * BF15
BF20 = max(0, 10 - RPTLAG ) * BF4
BF21 = ( CLT02 = 1)
BF23 = POLLAG * BF21
BF24 = ( ACC15 = 1) * BF16

Y = 0.580 - 0.174 * BF1 - 0.414 * BF3 + 0.196 * BF5 - 0.234 * BF7
+ 0.455 * BF9 + 0.131 * BF11 - 0.011 * BF15 - 0.006 * BF16 +
.135E-03 * BF18 - 0.013 * BF20 + .286E-03 * BF23 + 0.010 * BF24
```

[9] Because of the strong relationship between a missing value on treatment lag and the dependent variable, and the high percentage of claims in this category which had emergency room visits, an indicator variable for emergency room visits was tested as a surrogate. It was found not to be significant.

Table 10

Description of Categorical Variables

Variable	Value	Description
LEGALREP	1	No legal representation
	2	Has legal representation
INJ01	1	Injury consisted of strain or sprain only
	2	Injury did not consist of strain or sprain only
ACC04	1	Single vehicle accident
	2	Two or more vehicle accident
ACC14	1	Property damage was inconsistent with accident
	2	Property damage was consistent with accident
PARTDIS	1	Partially disabled
	2	Not partially disabled
CLT02	1	Had a history of previous claims
	2	No history of previous claims
ACC15	1	Was very minor impact collision
	2	Was not very minor impact collision

The basis functions and regression produced by MARS assist the analyst in understanding the impact of the predictor variables on the dependent variable. From the formulae above, it can be concluded that

1) when a lawyer is not involved (LEGALREP = 1), the probability of fraud or abuse declines by about 0.17
2) when the claimant has legal representation and the injury is consistent with a sprain or strain only, the probability of fraud or abuse increases by 0.2
3) when the claimant does not receive treatment from an outpatient health care provider (TRTLAG = missing), the probability of abuse declines by 0.41
4) a single vehicle accident where the claimant receives treatment from an outpatient health care provider (treatment lag not missing) decreases the probability of fraud by 0.23
5) if property damage is inconsistent with the accident, the probability of fraud or abuse increases by 0.46
6) if the claimant is partially disabled and receives treatment from an outpatient health care provider the probably of fraud or abuse is increased by 0.13

Of the red flag variables, small contributions were made by the claimant having a previous history of a claim[10] and the accident being a minor impact collision. Of the objective continuous variables obtained from the claim file, variables such as claimant age, report lag and policy lag have a small impact on predicting fraud or abuse.

Figures 11 and 12 display how MARS modeled the impact of selected continuous variables on the probability of fraud and abuse. For claims receiving outpatient health

[10] This variable only captures history of a prior claim if it was recorded by the insurance company. For some companies participating in the study, it was not recorded.

care, report lag has a positive impact on the probability of abuse, but its impact reaches its maximum value at about 10 days. Note the interaction between claimant age and ambulance costs displayed in Figure 12. For low ambulance costs, the probability of abuse rises steeply with claimant age and maintains a relatively high probability except for the very young and very old claimants. As ambulance costs increase, the probability of fraud or abuse decreases, and the decrease is more pronounced at lower and higher ages. Ambulance cost appears to be acting as a surrogate for injury severity.

Figure 11

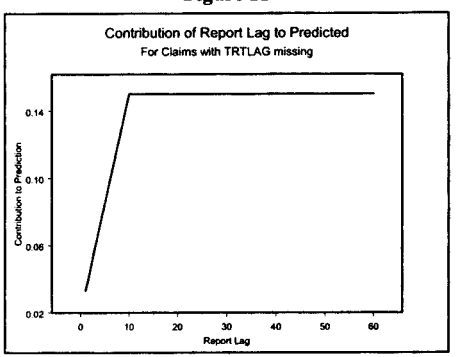
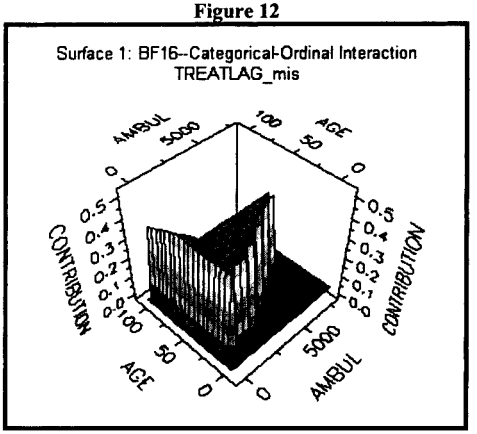


Figure 12



This section on explaining the model illustrates one of the very useful qualities of MARS as compared to neural networks: the output of the model is a formula which describes the relationships between predictor and dependent variables and which can be used to explain the model to management. To some extent, the sensitivity measure assists us in understanding the relationships fit by the neural network model, as it provides a way to assess the importance of each of the variables to the prediction. However, the actual functional relationships between independent and dependent variables are not typically available and the model can be difficult to explain to management.[11]

Evaluating the Goodness of the Fit and Comparing the Accuracy

One approach for testing the accuracy of models that is commonly used in data mining applications is to have separate training and testing samples. This approach was used in the previous example. Typically one half to one third of the data is held out for testing. However, when the database used for modeling is small, the analyst may not want to lose a large portion of the data to testing. Moreover, as the testing is performed on a relatively small sample, the goodness of fit results may be sensitive to random variation in the subsets selected for training and testing. An alternative procedure that allows more of the data to be used for fitting and testing is cross-validation. Cross-validation involves iteratively holding out part of the sample, fitting the model to the remainder of the sample and testing the goodness of the fitted model on the held out portion. For instance, the sample may be divided into 4 groups. Three of the groups are used to fit the model and one is used for testing. The process is repeated four times, and the goodness of fit statistics for the four test samples are averaged. As the AIB database is relatively small for a data mining application, this is the procedure used. Testing was performed using four fold cross-validation.

Both a MARS model and a neural network model were fit to four samples of the data. Each time the fitted model was used to predict the probability of fraud or abuse for one quarter of the data that was held out. The predictions from the four test samples were then combined to allow comparison of the MARS and neural network procedures.

Table 11 presents some results of the analysis. This table presents the R^2 of the regression of ASSESS on the predicted value from the model. The table shows that the neural network R^2 was higher than that of MARS. The table also displays the percentage of observations whose values were correctly predicted by the model. The predictions are based only on the samples of test claims. The neural network model correctly predicted 79% of the test claims, while MARS correctly predicted 77% of the test claims.

Table 11

Four Fold Cross-validation

Technique	R^2	Percent Correct
MARS	0.35	0.77
Neural Network	0.39	0.79

[11] Plate (2000) and Francis (2001) present a method to visualize the relationships between independent and dependent variables, The technique is not usually available in data mining software.

Tables 12 and 13 display the accuracy of MARS and the neural network in classifying fraud and abuse claims.[12] A cutoff point of 50% was used for the classification. That is, if the model's predicted probability of a 1 on ASSESS exceeded 50%, the claim was deemed an abuse claim. Thus, those claims in cell Actual =1 and Predicted=1 are the claims assessed by experts as probably abusive which were predicted to be abusive. Those claims in cell Actual=1, Predicted =0, are the claims assessed as probable abuse claims which were predicted by the model to be legitimate.

Table 12

MARS Predicted * Actual			
Predicted	**Actual**		
	0	1	Total
0	738	160	898
1	157	344	501
Total	895	505	

Table 13

Neural Network Predicted * Actual			
Predicted	**Actual**		
	0	1	Total
0	746	127	873
1	149	377	526
Total	895	505	

Table 14 presents the sensitivity and specificity of each of the models. The sensitivity is the percentage of events (in this case suspected abuse claims) that were predicted to be events. The specificity is the percentage of nonevents (in this case claims believed to be legitimate) that were predicted to be nonevents. Both of these statistics should be high for a good model. The table indicates that both the MARS and neural network models were more accurate in predicting nonevent or legitimate claims. The neural network model had a higher sensitivity than the MARS model, but both were approximately equal in their specificities. The neural network's higher overall accuracy appears to be a result of its greater accuracy in predicting the suspected fraud and abuse claims. Note that the sensitivity and specificity measures are dependent on the choice of a cutoff value. Thus, if a cutoff lower than 50% were selected, more abuse claims would be accurately predicted and fewer legitimate claims would be accurately predicted.

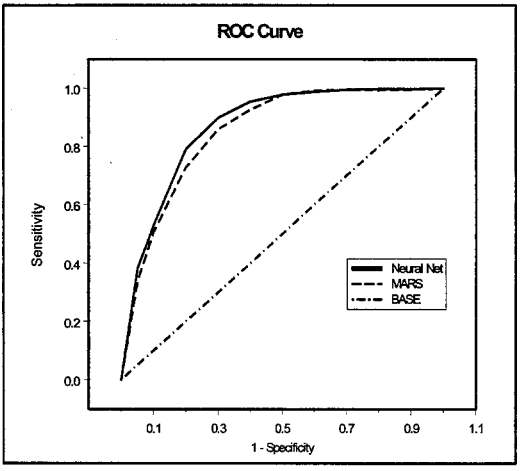
Table 14

Model	Sensitivity	Specificity
MARS	68.3	82.5
Neural Network	74.8	83.4

[12] These tables are often referred to as confusion matrices

A common procedure for visualizing the accuracy of models used for classification is the receiver operating characteristics (ROC) curve. This is a curve of sensitivity versus specificity (or more accurately 1.0 minus the specificity) over a range of cutoff points. When the cutoff point is very high (i.e. 1.0) all claims are classified as legitimate. The specificity is 100% (1.0 minus the specificity is 0), but the sensitivity is 0%. As the cutoff point is raised, the sensitivity increases, but so does 1.0 minus the specificity. Ultimately a point is reached where all claims are predicted to be events, and the specificity declines to zero. The baseline ROC curve (where no model is used) can be thought of as a straight line from the origin with a 45-degree angle. If the model's sensitivity increases faster than the specificity decreases, the curve "lifts" or rises above a 45-degree line quickly. The higher the "lift", the more accurate the model. It can be seen from the graph of the ROC curve that both the MARS and neural network models have significant "lift" but the neural network model has more "lift" than the MARS model.

Figure 13



A statistic that summarizes the predictive accuracy of a model as measured by an ROC curve is the area under the ROC curve (AUROC). A curve that rises quickly has more area under the ROC curve. Table 15 displays the AUROC for both models, along with their standard deviations and 95% confidence intervals. As the lower bound of the confidence interval for the neural network is below the higher bound of the confidence interval for MARS, it can be concluded that differences between the MARS model and the neural network model are not statistically significant.

Table 15

Statistics for Area Under the ROC Curve					
Test Result Variables	Area	Std Error	Asymptotic Sig	Lower 95% Bound	Upper 95% Bound
MARS Probability	0.85	0.01	0.000	0.834	0.873
Neural Probability	0.88	0.01	0.000	0.857	0.893

Summary of Comparison

The ROC curve results suggest that in this analysis the neural network enjoyed a modest though not statistically significant advantage over MARS in predictive accuracy. It should be noted that the database used for this study was quite small for a data mining application and may produce results that do not generalize to larger applications. Steinberg (Steinberg, 2001) reports that on other applications MARS equaled or exceeded the performance of neural networks. It should also be noted that some of the key comparative strengths of MARS such as its ability to handle missing data were not a significant factor in the analysis, as all but one of the variables were fully populated. [13] In addition, MARS's capability of clustering levels of categorical variables together was not relevant to this analysis, as no categorical variable had more than two levels.

A practical advantage that MARS enjoys over neural networks is the ease with which results can be explained to management. Thus, one potential use for MARS is to fit a model using neural networks and then apply MARS to the fitted values to understand the functional relationships fitted by the neural network model. The results of such an exercise are shown below:

BF1 = (LEGALREP = 1)
BF2 = (LEGALREP = 2)
BF3 = (TRTLAG ≠ missing)
BF4 = (TRTLAG = missing)
BF5 = (INJ01 = 1)
BF7 = (ACC04 = 1) * BF3
BF8 = (ACC04 = 2) * BF3
BF9 = (PARTDIS = 1) * BF8
BF11 = max(0, AMBUL - 182) * BF2
BF12 = max(0, 182 - AMBUL) * BF2
BF13 = (ACC14 = 1) * BF3
BF15 = (CLT02 = 1) * BF3
BF17 = max(0, POLLAG - 21) * BF3
BF19 = max(0, AGE - 41) * BF3
BF20 = max(0, 41 - AGE) * BF3

[13] One of the claims was missing data on the AGE variable, and this claim was eliminated from the neural network analysis and from comparisons of MARS the neural network model. Had more claims been missing the AGE variable, we would have modeled it in the neural network.

BF21 = (INS06 = 1)
BF23 = max(0, RPTLAG - 24) * BF8
BF24 = max(0, 24 - RPTLAG) * BF8
BF25 = BF1 * BF4
BF27 = (ACC15 = 1) * BF8
BF29 = (INJ03 = 1) * BF2

Y = 0.098 - 0.272 * BF1 + 0.334 * BF3 + 0.123 * BF5 - 0.205 * BF7 + 0.145 * BF9 - .623E-04 * BF11 + .455E-03 * BF12 + 0.258 * BF13 + 0.100 * BF15 + .364E-03 * BF17- 0.004 * BF19 - 0.001 * BF20 + 0.152 * BF21 + .945E-03 * BF23 - 0.002 * BF24 + 0.135 * BF25 + 0.076 * BF27 - 0.073 * BF29

This model had an R^2 of 0.9. Thus, it was able to explain most of the variability in the neural network fitted model. Though the sensitivity test revealed that LEGALREP is the most significant variable in the neural network model, its functional relationship to the probability of fraud is unknown using standard neural network modeling techniques. As interpreted by MARS, the absence of legal representation reduces the probability of fraud by 0.272., even without interacting with other variables. LEGALREP also interacts with the ambulance cost variable, INJ03 (police report shows no injury) and no use of a health care provider (treatment lag missing). The sensitivity measure indicated that the presence or absence of a value for treatment lag was the second most important variable. As stated earlier, this variable can be viewed as a surrogate for use of an outpatient health care provider. The use of an outpatient health care provider (TRTLAG $\neq$ missing) adds 0.334 to the probability of fraud or abuse, but this variable also interacts with the policy lag, report lag, claimant age, partial disability, ACC04, (single vehicle accident), ACC14 (property damage inconsistent with accident) and CLT02 (history of prior claims).

The MARS model helps the user understand not only the nonlinear relationships uncovered by the neural network model, but also describes the interactions which were fit by the neural network.

A procedure frequently used by data mining practitioners when two or more approaches are considered appropriate for an application is to construct a hybrid model or average the results of the modeling procedures. This approach has been reported to reduce the variance of the prediction (Salford Systems, 1999). Table 16 displays the AUROC statistics resulting from averaging the results of the MARS and neural network models. The table indicates that the performance of the hybrid model is about equal to the performance of the neural network. (The graph including the ROC curve for the combined model is not shown, as the curve is identical to Figure 13 because the neural network and combined curves cannot be distinguished.) Salford Systems (Salford Systems, 1999) reports that the accuracy of hybrid models often exceeds that of its components, but usually at least equals that of the best model. Thus, hybrid models that combine the results of two techniques may be preferred to single technique models because uncertainty about the accuracy of the predicted values on non-sample data is reduced.

Table 16

Statistics for Area Under the ROC Curve					
Test Result Variables	Area	Std Error	Asymptotic Sig	Lower 95% Bound	Upper 95% Bound
MARS Probability	0.853	0.01	0.000	0.834	0.873
Neural Probability	0.875	0.01	0.000	0.857	0.893
Combined Probability	0.874	0.01	0.000	0.857	0.892

Using Model Results

The examples in this paper have been used to explain the MARS technique and compare it to neural networks. The final example in this paper has been a fraud and abuse application that used information about the PIP claim that would typically be available shortly after the claim is reported to predict the likelihood that the claim is abusive or fraudulent. The results suggest that a small number of variables, say about a dozen, are effective in predicting fraud and abuse. Among the key variables in importance for both the neural network model and MARS are use of legal representation, use of an outpatient health care provider (as proxied by TRTLAG missing) and involvement in a single vehicle accident. Due to the importance of legal representation, it would appear useful for insurance companies to record information about legal representation in computer systems, as not all companies have this data available.

The results of both the MARS and neural network analysis suggest that both claim file variables (present in most claims databases) and red flag variables (common wisdom about which variables are associated with fraud) are useful predictors of fraud and abuse. However, this and other studies support the value of using analytical tools for identifying potentially abusive claims. As pointed out by Derrig (Derrig, 2002), fraud models can help insurers sort claims into categories related to the need for additional resources to settle the claim efficiently. For instance, claims assigned a low score by a fraud and abuse model, can be settled quickly with little investigative effort on the part of adjusters. Insurers may apply increasingly greater resources to claims with higher scores to acquire additional information about the claimant/policyholder/provider and mitigate the total cost of the claim. Thus, the use of a fraud model is not conceived as an all or nothing exercise that classifies a claim as fraudulent or legitimate, but a graduated effort of applying increasing resources to claims where there appears to be a higher likelihood of material financial benefit from the expenditures.

Conclusion

This paper has introduced the MARS technique and compared it to neural networks. Each technique has advantages and disadvantages and the needs of a particular application will determine which technique is most appropriate.

One of the strengths of neural networks is their ability to model highly nonlinear data. MARS was shown to produce results similar to neural networks in modeling a nonlinear function. MARS was also shown to be effective at modeling interactions, another strength of neural networks.

In dealing with nominal level variables, MARS is able to cluster together the categories of the variables that have similar effects on the dependent variable. This is a capability not possessed by neural networks that is extremely useful when the data contain categorical variables with many levels such as ICD9 code.

MARS has automated capabilities for handling missing data, a common feature of large databases. Though missing data can be modeled with neural networks using indicator variables, automated procedures for creating such variables are not available in most standard commercial software for fitting neural networks. Moreover, since MARS can create interaction variables from missing variable basis functions and other variables, it can create surrogates for the missing variables. Thus, on applications using data with missing values on many variables, or data where the categorical variables have many values, one may want to at least preprocess the data with MARS to create basis functions for the missing data and categorical variables which can be used in other procedures.

A significant disadvantage of neural networks is that they are a "black box". The functions fit by neural networks are difficult for the analyst to understand and difficult to explain to management. One of the very useful features of MARS is that it produces a regression like function that can be used to understand and explain the model; therefore it may be preferred to neural networks when ease of explanation rather than predictive accuracy is required. MARS can also be used to understand the relationships fit by other models. In one example in this paper MARS was applied to the values fit by a neural network to uncover the important functional relationships modeled by the neural network.

Neural networks are often selected for applications because of their predictive accuracy. In a fraud modeling application examined in this paper the neural network outperformed MARS, though the results were not statistically significant. The results were obtained on a relatively small database and may not generalize to other databases. In addition, the work of other researchers suggests that MARS performs well compared to neural networks. However, neural networks are highly regarded for their predictive capabilities. When predictive accuracy is a key concern, the analyst may choose neural networks rather than MARS when neural networks significantly outperform MARS. An alternative approach that has been shown to improve predictive accuracy is to combine the results of two techniques, such as MARS and neural networks, into a hybrid model.

This analysis and those of other researchers supports the use of intelligent techniques for modeling fraud and abuse. The use of an analytical approach can improve the performance of fraud detection procedures that utilize red flag variables or subjective claim department rules by 1) determining which variables are really important in predicting fraud, 2) assigning an appropriate weight to the variables when using them to predict fraud or abuse, and 3) using the claim file and red flag variables in a consistent manner across adjusters and claims.

References

Allison, Paul. *Missing Data*, Sage Publications, 2001

Berry, Michael J. A., and Linoff, Gordon, *Data Mining Techniques*, John Wiley and Sons, 1997

Brockett, Patrick L., Xiaohua Xia and Richard A. Derrig, 1998, "Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud", *Journal of Risk and Insurance*, June, 65:2.

Brockett, Patrick L., Richard A. Derrig, Linda L. Golden, Arnold Levine and Mark Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs", *Journal of Risk and Insurance*, September, 2002, pp. 341-371.

Dhar, Vasant and Stein, Roger, *Seven Methods for Transforming Corporate Data Into Business Intelligence*, Prentice Hall, 1997

Derrig, Richard A., Herbert I. Weisberg and Xiu Chen, 1994, "Behavioral Factors and Lotteries Under No-Fault with a Monetary Threshold: A Study of Massachusetts Automobile Claims", *Journal of Risk and Insurance*, June, 1994, 61:2: 245-275.

Derrig, Richard A., and Krzysztof M. Ostaszewski, "Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification", *Journal of Risk and Insurance*, September, 1995, 62:3: 447-482.

Derrig, Richard, "Patterns, Fighting Fraud With Data", *Contingencies*, Sept/Oct, 1999, pp. 40–49.

Derrig, Richard A., and Valerie Zicko, "Prosecuting Insurance Fraud: A Case Study of The Massachusetts Experience in the 1990s", Risk Management and Insurance Research, 2002.

Derrig, Richard, "Insurance Fraud", *Journal of Risk and Insurance*, September, 2002, pp. 271-287

Francis, Louise, "Neural Networks Demystified", *Casualty Actuarial Society Forum,* Winter 2001, *pp 253 - 320.*

Freedman, Roy S., Klein, Robert A. and Lederman, Jess, *Artificial Intelligence in the Capital Markets*, Probus Publishers 1995

Hastie, Trevor, Tibshirani, Robert, *Generalized Additive Models*, Chapman and Hall, 1990

Hastie, Trevor, Tibshirani, Robert and Freidman, Jerome, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2001

Hayward, Gregory, "Mining Insurance Data to Promote Traffic Safety and Better Match Rates to Risk", *Casualty Actuarial Society Forum,* Winter 2002, *pp. 31 – 56.*

Hosmer, David W. and Lemshow, Stanley, *Applied Logistic Regression*, John Wiley and Sons, 1989

Keefer, James, "Finding Causal Relationships By Combining Knowledge and Data in Data Mining Applications", Paper presented at Seminar on Data Mining, University of Delaware, April, 2000.

Lawrence, Jeannette, *Introduction to Neural Networks: Design, Theory and Applications*, California Scientific Software, 1994

Little, Roderick and Rubin, Donald, *Statistical Analysis with Missing Data*, John Wiley and Sons, 1987

Marsh, Lawrence and Cormier, David, *Spline Regression Models*, Sage Publications, 2002

Martin, E. B. and Morris A. J., "Artificial Neural Networks and Multivariate Statistics", in *Statistics and Neural Networks: Advances at the Interface,* Oxford University Press, 1999, pp. 195 – 292

Miller, Robert and Wichern, Dean, *Intermediate Business Statistics*, Holt, Reinhart and Winston, 1977

Neal, Bradford, *Bayesian Learning for Neural Networks*, Springer-Verlag, 1996

Plate, Tony A., Bert, Joel, and Band, Pierre, "Visualizing the Function Computed by a Feedforward Neural Network", *Neural Computation*, June 2000, pp. 1337-1353.

Potts, William J.E., *Neural Network Modeling: Course Notes*, SAS Institute, 2000

Salford Systems, "Data mining with Decision Trees: Advanced CART Techniques", Notes from Course, 1999

Smith, Murry, *Neural Networks for Statistical Modeling*, International Thompson Computer Press, 1996

Speights, David B, Brodsky, Joel B., Chudova, Durya L., "Using Neural Networks to Predict Claim Duration in the Presence of Right Censoring and Covariates", *Casualty Actuarial Society Forum*, Winter 1999, pp. 255-278.

Steinberg, Dan, "An Introduction to MARS", Salford Systems, 1999

Steinberg, Dan, "An Alternative to Neural Networks: Multivariate Adaptive Regression Splines (MARS)", *PC AI*, January/February, 2001, pp. 38 – 41

Venebles, W.N. and Ripley, B.D., Modern *Applied Statistics with S-PLUS*, third edition, Springer, 1999

Viaene, Stijn, Derrig, Richard, Baesens, Bart and Dedene Guido, "A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Fraud Detection", *Journal of Risk and Insurance*, September, 2002, pp. 373 – 421

Warner, Brad and Misra, Manavendra, "Understanding Neural Networks as Statistical Tools", *American Statistician*, November 1996, pp. 284 – 293

Weisberg, Herbert I., and Richard A. Derrig, "Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Claims in Massachusetts", *Journal of Insurance Regulation*, June, 1991, 9:4: 497-541.

Weisberg, Herbert I., and Richard A. Derrig, "Quantitative Methods for Detecting Fraudulent Automobile Injury Claims", AIB Cost Containment/Fraud Filing (DOI Docket R95-12), Automobile Insurers Bureau of Massachusetts, July,1993, 49-82.

Weisberg, Herbert I., and Richard A. Derrig, "Identification and Investigation of Suspicious Claims", AIB Cost Containment/Fraud Filing (DOI Docket R95-12), Automobile Insurers Bureau of Massachusetts, July, 1995, 192-245, Boston.

Weisberg, Herbert I., and Richard A. Derrig, "Massachusetts Automobile Bodily Injury Tort Reform", *Journal of Insurance Regulation*, Spring, 1992, 10:3:384-440.

Zapranis, Achilleas and Refenes, Apostolos-Paul, *Principles of Neural Network Model Identification Selection and Adequacy*, Springer-Verlag, 1999