2·0·1·6

Exam 9 Study Kit

CAS 2016 STUDY KIT

Exam 9
Financial Risk and Rate of Return

1. Butsic, R.P., "Solvency Measurement for Property-Liability Risk-Based Capital Applications," *The Journal of Risk and Insurance*, American Risk and Insurance Association, Inc., Volume 61, No. 4 (December 1994), pp. 656-690.

2. Coval, J.; Jurek, J.; and Stafford, E., "The Economics of Structured Finance," *Journal of Economic Perspectives*, American Economic Association, Volume 23, Number 1, Winter 2009.

3. Culp, C.L.; Miller, M.H.; and Neves, A.M.P., "Value at Risk: Uses and Abuses," *The New Corporate Finance: Where Theory Meets Practice* (Third Edition), Chew, D.H., editor; McGraw-Hill/Irwin, 2001, Chapter 33, pp. 462-471.

4. Cummins, J. D., "Allocation of Capital in the Insurance Industry", *Risk Management and Insurance Review*, American Risk and Insurance Association, Inc., Spring 2000, Vol. 3, No. 1, pp. 7-27.

5. Cummins, J. D., "CAT Bond and Other Risk-Linked Securities: State of the Market and Recent Developments," *Risk Management and Insurance Review*, American Risk and Insurance Association, Inc., 2008, Vol. 11, No. 1, pp. 23-47.

6. Noris, P.D., "Asset/Liability Management Strategies for Property and Casualty Companies," Morgan Stanley, May 1985.

7. Stulz, R.M., "Rethinking Risk Management," *The New Corporate Finance: Where Theory Meets Practice* (Third Edition), Chew, D.H., editor; McGraw-Hill/Irwin, 2001, Chapter 29, pp. 411-427.

CAS Exam 9 Study Kit

"Solvency Measurement for Property-Liability Risk-Based Capital Applications"

by R.P. Butsic

The Journal of Risk and Insurance

© The Journal of Risk and Insurance, 1994 Vol. 61, No. 4, 656-690

Solvency Measurement for Property-Liability Risk-Based Capital Applications

Robert P. Butsic

ABSTRACT

Regulators have recently adopted a risk-based capital formula for property-liability insurers. This article develops practical methods for setting risk-based capital standards using the expected policyholder deficit as the solvency measure. The analysis considers the stochastic nature of insurance risk, using market valuation to remove measurement bias, and finds that a proper time horizon is the period between risk-based capital evaluations. The present value of the expected policyholder deficit is shown to be equivalent to a financial option implicitly given by the policyholders. Finally, covariance of risk elements is considered, deriving a simple square root rule.

Introduction

The recent failure of several large life insurers, following the disastrous experience of the savings and loan industry, has pushed solvency oversight to the top of the regulatory agenda. In late 1990, the National Association of Insurance Commissioners began a mission to establish risk-based capital formulas for both life and property-liability insurance, as well as model laws to institute the capital requirements.

Formula-driven capital requirements are not new to insurance. For about 40 years, European authorities have used various formulas to set solvency margins (see Byrnes, 1986, or Daykin et al., 1987, for an extensive overview of international approaches to solvency regulation). In the United States, detailed risk-based capital formulas for other financial institutions (banks and savings and loans) have been adopted and are now undergoing a phase-in period.

This article shows how risk can be quantified in establishing risk-based capital for property-liability insurers. The first section discusses the roles of parties to the insurance contract, establishing why the risk-based capital concept is economically useful. The second section defines capital and risk, lead-

---

Robert P. Butsic is Assistant Vice President of the Fireman's Fund Insurance Companies. The author is indebted to David Cummins, Richard Derrig, and an anonymous referee for their helpful comments. The author is also grateful for the guidance of David Hartman and the other members of the American Academy of Actuaries P & C Risk-Based Capital Task Force. This committee's work inspired much of the author's efforts and provided a real-world application of the theory.

ing to the expected policyholder deficit as the relevant risk measure for solvency analysis. A balance sheet model relates capital levels to the size of the expected deficit, providing results for the commonly-used normal and lognormal distributions.

The third section introduces the time dimension with a discussion of bias. Market valuation, for both assets and liabilities, is used to remove bias in risk measurement. This section next describes diffusion processes to show how insurance risk is time-dependent and then shows that a proper time horizon for solvency determination is the period between risk-based capital evaluations. The model is completed by taking the present value of the policyholder deficit and showing that this measure is equivalent to a financial option implicitly given by the policyholders.

The fourth section shows that the degree of correlation between risk elements is a critical factor in properly setting capital levels. Using linear approximation, a simple square root rule incorporates the correlation; its application is illustrated with a hypothetical balance sheet. The final section discusses implications and applications of the results.

Economic Basis for Risk-Based Capital

The purpose of insurance is to spread the costs of unforeseen economic loss over a wide base of policyholders. In turn, the main purpose of solvency regulation is to ensure that the promised insurance protection is available to an acceptable degree of certainty.

The solvency of an insurer is intimately linked to the condition of its balance sheet. Capital, the excess of assets over obligations, represents the owners' stake, or equity in the firm. Under statutory accounting principles (SAP), capital is called surplus; under generally accepted accounting principles (GAAP), capital is called equity.

An insolvency occurs when obligations exceed assets, a condition called *technical insolvency.* Usually at this point regulators will have intervened to place the company in conservatorship or will have severely curtailed its operations (theoretically, however, an insurer could operate beyond the point of technical insolvency if payment of losses and expenses sufficiently lagged cash inflows). The capital providers lose their entire stake in the insurer, and the holders of the obligations, mostly policyholders, take over the assets.

In general, risk-based capital is the theoretical amount of capital needed to absorb the risks of conducting a business. For insurance, it is the amount of capital necessary to assure the major parties to an insolvency that the danger of failure is acceptably low. The standard for this low expectation will be addressed below.

For assessing the consequences of insolvency, the primary parties to the insurance contract are the policyholder and the equityholders—providers of capital to the insurance firm. Note that the policyholder can also be a capital supplier in the case of a mutual insurer. Both the insurer and the regulator are intermediaries. Third-party claimants also have a stake in the success of the insurer.

In a perfectly efficient market, solvency regulation would not be necessary. Consumers would know the likelihood of their insurer's becoming insolvent, with the price of the policy being adjusted to reflect the expectation that not all claims would be fully paid. Also, insurers could adjust their capital levels to reflect their customers' preferences for more or less protection at higher and lower prices. The result would likely be a range of capitalization from high to low leverage, with increasing degrees of policyholder security.

In the real world, however, solvency regulation exists despite the availability of considerable public information regarding the financial strength of insurers and a strong demand for solvency. Many policyholders, especially business customers, actively research their insurers' financial soundness. However, not all policyholders can efficiently obtain this information. Thus, regulators have determined that solvency protection is desirable, particularly for the less informed personal lines customers (although without the protection afforded by regulation, market mechanisms would arise to disseminate solvency information).[1]

Given the existence of solvency regulation, it is reasonable to assume that regulators should provide the public with a minimum level of protection from the adverse effects of insurer insolvency. Through a risk-based capital program, regulation can provide that minimum level, with additional security provided by the competitive insurance market.

Desirable Features of a Risk-Based Capital Method

In order to ensure equity for all parties to the insurance contract (policyholders, claimants, capital providers, and insurers) and to be practicable, the risk-based capital method should satisfy several criteria. First, the solvency standard should be the same for all classes of the above parties (e.g., personal vs. commercial insureds, second- vs. third-party claimants, and primary insurers vs. reinsurers).[2]

Second, the risk-based capital (RBC) should be objectively determined; that is, two insurers with the same risk measures should have identical risk-based capital. Also, a single insurer should obtain the same results under different

---

[1] A reviewer has argued that the "consumer ignorance" hypothesis as a basis for regulation is not compelling. First, guaranty funds offer limited protection, requiring insurers to provide more private capital. Second, firms place a high value on ratings from private organizations like Standard & Poor's; there is currently a trend toward production of information regarding financial strength. Third, consumers are not considered ignorant when purchasing corporate securities, which entail risks similar to insurance failure.

[2] There is an issue as to whether the policyholder, in view of the price paid for the policy, actually anticipated potential nonpayment of claims. Some would argue that a customer accepting the contract for a low price has implicitly self-insured, with the premium saved being the expected value of the shortage. This objection could be partially met by establishing different insolvency standards for the different classes. However, since an insurer's insolvency affects *all* of its policyholders, this scheme would be impractical, requiring companies to insure only one class.

regulatory jurisdictions using the same RBC method. The objectivity criterion dictates that the risk-based capital method can be expressed as a mathematical formula incorporating financial data from insurers.

Third, the method must be able to discriminate between quantifiable items of meaningful value that differ materially in their riskiness. For example, if stocks are significantly riskier than bonds, and the amounts of these two assets are known for each insurer, then the risk-based capital method should incorporate the distinction. Each such distinct item is defined as a *risk element*. As shown below, when discussing the effect of time, a risk element must be a balance sheet quantity.

These features will be useful in the development of appropriate solvency measures for a risk-based capital program. The goal of this article—to determine how much capital is needed for the entire insurer—is accomplished by evaluating each risk element singly and then combining the capital amounts of all risk elements.

Expected Deficit as a Measure of Insolvency Risk

In solvency analysis, the relevant risk is that obligations (primarily reserves) may exceed assets, both items being balance sheet quantities. For a balance sheet item, risk is present when the future realization of the item can be one of several values, but the particular outcome is currently unknown.

To clarify the discussion, I use a simplified model with a parallel numerical example; both will be extended to incorporate additional features.

The value of assets is denoted by $\tilde{A}$; assets are cash (the realizable value is certain). The loss reserve $\tilde{L}$ is the unpaid loss, whose realizable value is a random variable. The capital $\tilde{C}$ equals assets minus the loss reserve. The realizable value of capital is a random variable.[3]

For simplicity, we initially assume that the passage of time does not affect value (i.e., interest is zero), other assets and liabilities are zero (e.g., receivables and tax liability), there are no other transactions (taxes, expenses, etc.), losses include loss adjustment expenses, and losses are valued at the beginning of the year and paid at the end of the year.

To determine how much capital is needed to provide the minimum security standard mentioned earlier, we must define the level of protection. The usual measure of risk with respect to insurance solvency is the *probability of ruin*. Although this measure may appear reasonable from the internal perspective of insurance management (whose employment opportunities depend more on the fact of insolvency than on its degree), it is inadequate for public policy.[4]

---

[3] Random variables are denoted by tildes; their expectations are denoted by plain type. Notice that since the loss is a random variable, the capital is also a random variable.

[4] Classical risk theory, which has guided the development of European solvency margins (e.g., Beard, Pentikäinen, and Pesonen, 1984) has ignored the severity of ruin. Even the simulation modeling by Daykin et al. (1987), which provides an extensive individual insurer approach to risk-based capital, casts its results in terms of ruin probabilities. A possible reason why so little

To illustrate, suppose that two insurers each have the same beginning balance sheets. Assets of \$13,000 equal unpaid losses of \$10,000 plus capital of \$3,000. However, their unpaid losses have different probability distributions, producing three possible end-of-year results for each insurer (see Table 1). The payoff to policyholders is limited to the insurer's assets of \$13,000. Both insurers have a 20 percent chance of becoming insolvent under Scenario 3, but Insurer B's policyholders are clearly worse off. They will on average forfeit 0.2(18,000 − 13,000) = \$1,000 of their claim payments. Insurer A's policyholders, on the other hand, will forego an expected 0.2(13,100 − 13,000) = \$20 of their claim payments. Clearly, the probability-of-ruin criterion is inadequate to express the policyholders' exposure to loss. It is not sufficient merely to consider the probability of ruin—its *severity* must also be appreciated.

Table 1
Two Insurers with Same Balance Sheet but Different Unpaid Loss Distribution—Asset Amount Is Certain

	Asset Amount	*Loss Amount*	*Capital Amount*	*Probability*	*Claim Payment*	*Deficit*
Insurer A						
Scenario 1	13,000	6,900		0.2	6,900	0
Scenario 2	13,000	10,000		0.6	10,000	0
Scenario 3	13,000	13,100		0.2	13,000	100
Expectation	13,000	10,000	3,000		9,980	20
Insurer B						
Scenario 1	13,000	2,000		0.2	2,000	0
Scenario 2	13,000	10,000		0.6	10,000	0
Scenario 3	13,000	18,000		0.2	13,000	5,000
Expectation	13,000	10,000	3,000		9,000	1,000

This example suggests that a reasonable measure of insolvency risk is the expected value of the difference between the amount the insurer is obligated to pay the claimant and the actual amount paid by the insurer.[5] We will call this difference the *policyholder deficit*.

Using the expected policyholder deficit (EPD) risk measure, we can consistently measure insolvency risk in such a way that a standard minimum level

attention has been given to ruin severity in the insurance literature is that the ruin concept originated with the study of gambler's ruin (by Bernoulli and others in the eighteenth century). In the classical gambling problem, one could not bet more than one's stake, so if ruin occurred, the gambler could not default. With insurance, however, the management can risk *more* than the owners' stake (capital) and so the depth of ruin is important.

[5] Here we ignore guaranty fund or other external sources of recoupment such as another insurer or the federal government. However, the value of insolvency insurance can be determined by methods outlined in this article. Cummins (1988) develops risk-based guaranty fund premiums using similar principles. Note that a more complete measure of insolvency cost is the *present value* of the deficit, a feature added below.

of protection is applied to all classes of policyholders and insurers. The EPD measure can apply equally to all risk elements, whether assets or liabilities. To adjust to the scale of different risk element sizes, the ratio of the expected policyholder deficit to expected loss—or the *EPD ratio*—is used as the basic measure of policyholder security. The EPD ratio is denoted by d. The respective EPD ratios for Insurers A and B are 0.002 and 0.100.

For a discrete loss size probability distribution, when assets are certain, the expected policyholder deficit is

$$D_L = \sum_{x>A} p(x)(x-A). \quad (1)$$

where $p(\bullet)$ = the probability density for losses ($0 \leq x < \infty$). The EPD ratio is $d_L = D_L/L$.

Expected Policyholder Deficit with Asset Risk

We now extend the preceding numerical example to risky assets (see Table 2). Insurer C has a known loss of $5,000 about to be paid, but it has $6,300 of assets whose year-end value is uncertain (for this example it is assumed that the expected future value of the assets equals their current value). Here the policyholders will come up short 10 percent of the time, when assets turn out to be worth $3,000. The deficit in this case is $2,000, giving an expected policyholder deficit of $200 and an EPD ratio of 0.04. This protection level occurs with a 0.206 ratio of capital to assets.

Table 2
Insurer with Asset Risk—Unpaid Loss Amount Is Certain

	Asset Amount	*Loss Amount*	*Capital Amount*	*Probability*	*Claim Payment*	*Deficit*
Scenario 1	12,000	5,000		0.1	5,000	0
Scenario 2	6,000	5,000		0.8	5,000	0
Scenario 3	3,000	5,000		0.1	3,000	2,000
Expectation	6,300	5,000	1,300		4,800	200

For a discrete probability distribution of asset values, where losses are certain, the EPD is the expectation of assets being less than losses:

$$D_A = \sum_{L>y} q(y)(L-y), \quad (2)$$

where $q(\bullet)$ = the probability density for the asset value ($0 \leq y < \infty$).

Setting Capital to a Common EPD Ratio

Now suppose that the regulator wishes to set a capital standard so that the expected policyholder deficit is the same for all insurers, at 5 percent of the expected losses. This amounts to $500 for Insurers A and B and $250 for C. In order to satisfy this requirement, it is assumed that the amount of the expected loss is given for each insurer, and the level of beginning assets is adjusted to reach the desired capital. The added or subtracted assets are the same type as the original assets, so that the probability distribution of ending asset values, relative to the beginning assets, stays constant. This homogeneity condition ensures that, like losses, the assets have the same risk (per unit) regardless of the amount of capital.[6]

The capital for Insurer B must be increased in order to meet the 5 percent expected policyholder deficit mark, while capital for A and C must be reduced to reach this standard. Table 3 summarizes this calculation, using Scenario 3 (the two favorable scenarios do not contribute to a policyholder deficit).

Table 3
Balance Sheets and Capital Requirements for Three Hypothetical Insurers with Assets Adjusted to Produce 5 Percent Expected Policyholder Deficit Ratio

	Insurer A	*Insurer B*	*Insurer C*
Scenario 3 Probability	0.2	0.2	0.1
Beginning Assets	10,600	15,500	5,250
Ending Assets	10,600	15,500	2,500
Expected Loss	10,000	10,000	5,000
Capital	600	5,500	250
Loss Amount	13,100	18,000	5,000
Claim Payment	10,600	15,500	2,500
Deficit	2,500	2,500	2,500
Expected Deficit	500	500	250
Capital/Expected Loss	0.060	0.550	
Capital/Assets			0.048

Note: The original beginning assets for Insurer C are \$6,300 with the Scenario 3 ending assets at 10/21 of this amount, or \$3,000. When the assets are reduced to produce the 5 percent expected policyholder deficit ratio, the 10/21 proportion is maintained, giving the above \$2,500 Scenario 3 ending assets.

Using the EPD with Continuous Probability Distributions

For a general discrete distribution of loss or asset values, there is no simple formula that will directly produce the amount of capital required to reach a given expected policyholder deficit value. One must iterate by varying the beginning assets. For continuous distributions, which better represent the distribution of actual realized balance sheet items, it is possible to obtain an explicit relationship between the EPD and capital.

Assume that assets are certain and liabilities (losses) are uncertain. From our earlier notation, we have $\tilde{A} = \tilde{L} + \tilde{C}$. Let $c = C/L$ be the capital per unit of

[6] This requirement allows the risk-based capital formula to apply a constant factor to an asset having a given risk. If, for example, to reduce a given EPD ratio, the added assets were riskless, then the asset risk per unit of assets would decline. An alternative to adding/subtracting assets is reducing/increasing the losses. This procedure will produce a different size for the balance sheet, but the assets, losses, and capital will be in the same proportions as when assets are modified.

expected loss. The expected loss is $L = \int_0^\infty xp(x)dx$, where $p(\bullet)$ is the probability density for the losses, and $0 \le x < \infty$.

The expected policyholder deficit here is analogous to the discrete version. It is the expectation of losses exceeding assets, or

$$D_L = \int_A^\infty (x-A)p(x)dx. \tag{3a}$$

Notice the similarity of this expression to the cost of excess loss coverage (x denotes the size of loss and A the retention). The excess loss cost is the sum of each possible loss amount exceeding the retention, weighted by the probability of occurrence.

For certain losses and uncertain assets, the expected policyholder deficit is also similar to its discrete counterpart. It is the expectation of assets being less than losses:

$$D_A = \int_0^L (L-y)q(y)dy, \tag{3b}$$

where $q(\bullet)$ = the probability density for the assets ($0 \le y < \infty$). From the balance sheet definition, the expected value of assets is $A = (1+c)L$.

The Appendix applies these formulas to derive the EPD ratios for the important case of normally distributed risk elements. Equations (4a) and (4b) express the results in terms of the ratio of capital to the expected value of the risk element, and its coefficient of variation (CV), or the ratio of the standard deviation of the risk element to its mean.

$$d_L = \frac{D_L}{L} = k\phi\left(\frac{-c}{K}\right) - c\Phi\left(\frac{-c}{k}\right) \tag{4a}$$

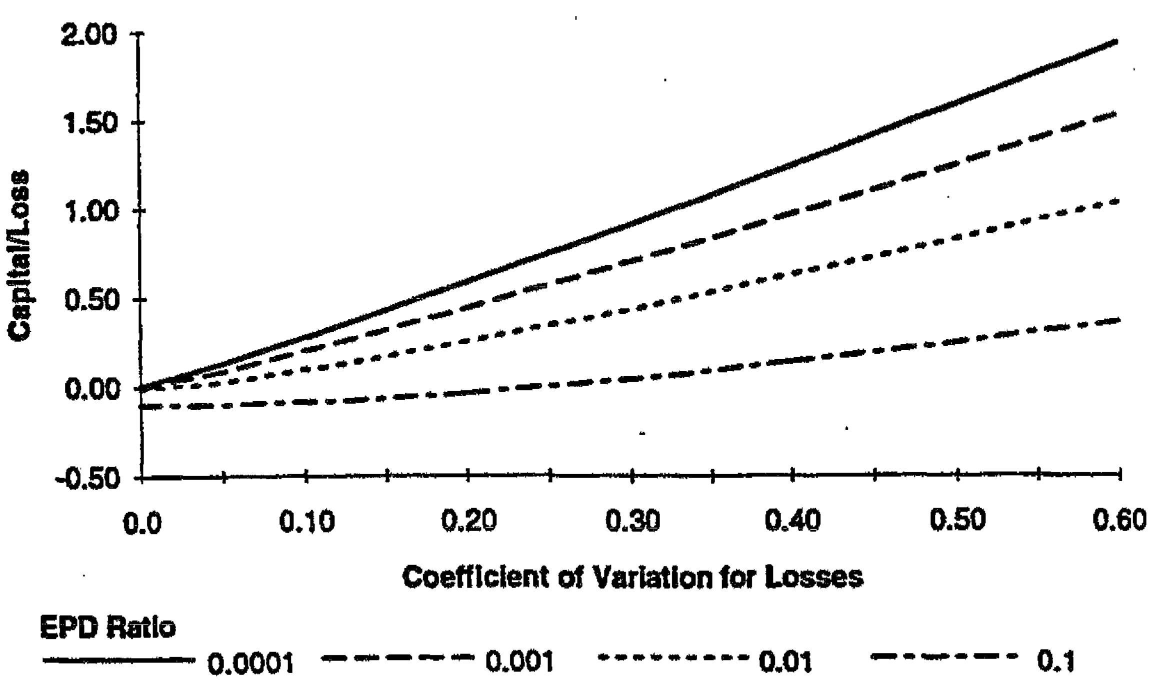
$$d_A = \frac{D_A}{L} = \frac{1}{1-c_A}\left[k_A\phi\left(\frac{-c_A}{k_A}\right) - c_A\Phi\left(\frac{-c_A}{k_A}\right)\right], \tag{4b}$$

where k = the CV of losses,
k_A = the CV of assets,
c_A = the capital/assets ratio,
$\Phi(\bullet)$ = the cumulative standard normal distribution, and
$\phi(\bullet)$ = the standard normal density function.

For the same coefficient of variation, in order for d_A to equal d_L, more capital (per unit of assets) is required for assets than for losses (per unit of losses). This is easily seen by setting $k_A = k$ and $c_A = c$ in equation (4b) and equating with (4a). We get $d_A = d_L/(1-c) > d_L$, for $c > 0$ (negative capital is discussed later). Since the EPD ratio is higher for assets than for losses, the asset capital ratio must be greater than c. Intuitively, asset risk under the normal distribution needs more capital than loss risk because (assuming the same beginning balance sheet), if assets and losses have the same CV, the standard deviation of assets is larger because the value of assets exceeds that of losses. Since the normal probability density is symmetric, the value of capital for both risky assets and losses is distributed with the same mean and with a larger standard deviation under asset risk.

Figure 1 illustrates, for losses as a risk element, capital/loss (c) values for a range of d_L and k values. For a particular EPD ratio, c can be well-approximated by a linear function of k. Notice that the capital ratio is *negative* for high values of d_L and low values of k. When k equals zero, there is *no risk* to policyholders—in order for there to be a positive policyholder deficit, the amount of (riskless) assets must necessarily be less than that of the certain loss. This situation, unlikely in practice, would guarantee a deficit equal to D_L.

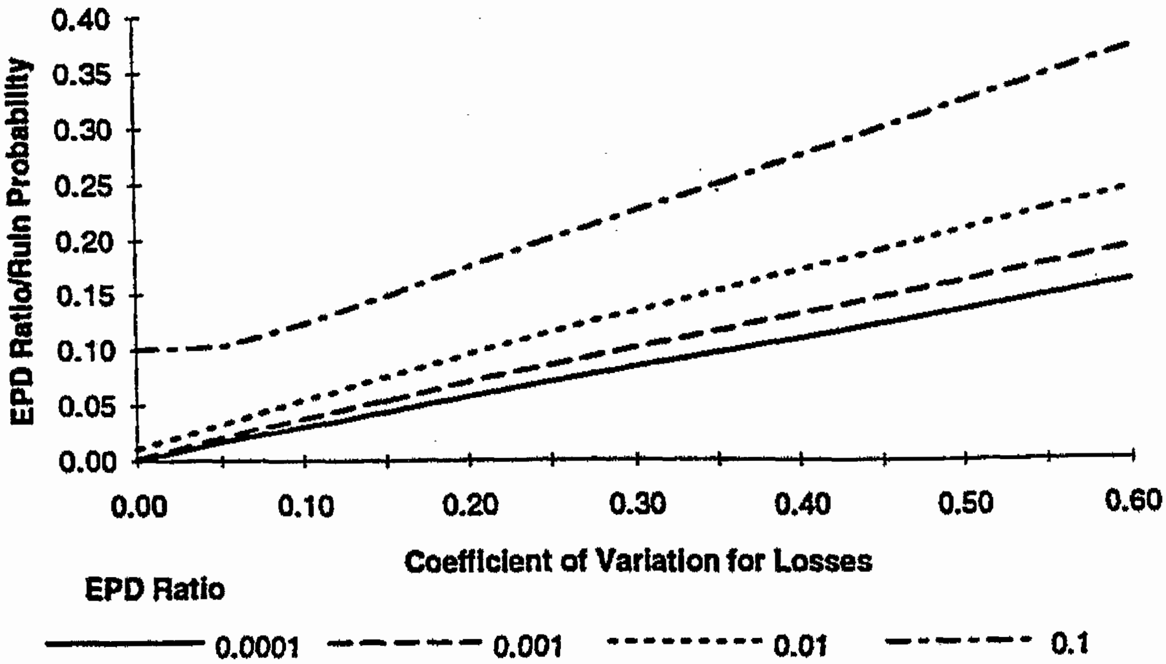
Figure 1
Capital/Loss vs. Expected Policyholder Deficit Ratio and Coefficient of Variation of Losses Under Normal Distribution



It is informative to compare the expected policyholder deficit ratio solvency criterion with the corresponding ruin probability. For loss and asset risk elements, the respective probabilities of ruin are simply $\Phi(-c/k)$ and $\Phi(-c_A/k_A)$. Figure 2 demonstrates that, under the normal distribution, the EPD ratio is an increasing function of the coefficient of variation and the ruin probability.[7] Thus, there is no single ruin probability corresponding to a given EPD ratio.

[7] This relationship follows directly from the comparative statics of the option model, whose analogy with the expected policyholder deficit is discussed below. Specifically, it is well known that an increase in risk of the underlying asset increases the value of a put option on that asset.

Figure 2
Expected Policyholder Deficit Ratio/Ruin Probability vs.
Coefficient of Variation of Losses Under Normal Distribution



The normal distribution might be a reasonable approximation for the variation of aggregate incurred loss amounts arising from a population having a known mean, where individual losses occur independently of each other. An example is noncatastrophe property insurance. For correlated events, and where the mean is unknown, a popular assumption is the *lognormal* distribution (see Aitchison and Brown, 1969, for a thorough explanation of the lognormal distribution and its economic applications). This has the desirable property that negative values are impossible, and the skewness of outcomes appears to accord with observed results. However, the sum of two lognormal variables is only approximately lognormally distributed (the product is a lognormal variable).[8]

For the lognormal distribution, the Appendix also derives analogous formulas for d_L and d_A. Although equation (3) can be used directly, the derivation uses option pricing methods to highlight the option properties of the EPD.[9] The capital ratios c and c_A are determined by solving

[8] Samuelson (1983, p. 556) proves that the sum of two lognormal variables is in fact lognormal, provided that the sum occurs in a continuous-time framework (as in the Black-Scholes, 1973, option model). However, here we are using a discrete time model.

[9] The standard model for pricing stock options assumes that the stock price exhibits instantaneous geometric Brownian motion, which implies that, at the end of any finite time period, the stock price has a lognormal distribution. The option price can be determined by taking the present value (at riskless interest) of the expected difference between the stock price and the exercise price, where the lognormal distribution has been transformed using a risk-neutral probability measure. This expectation is analogous to the EPD.

$$d_L = \Phi(a) - (1+c)\Phi(a-k) \tag{5a}$$

and

$$d_A = \Phi(b) - \frac{\Phi(b-k_A)}{1-c_A}, \tag{5b}$$

where $a = (k/2) - (\ln(1+c)/k)$,
$b = (k_A/2) + (\ln(1-c_A)/k_A)$, and
$\Phi(\bullet)$ = the cumulative normal distribution.

Unlike the normal distribution, when assets and losses have the same coefficient of variation, under the lognormal model d_A will equal d_L with a *smaller* capital ratio for assets than for losses. Setting $k_A = k$ in equation (5b) and equating with (5a), we get $c_A = c/(1+c) < c$. While the normal distribution is symmetric, the lognormal density is not: the policyholder deficit for losses (actual loss minus fixed assets) can be infinite, but for assets it is limited to the amount of the riskless loss (the minimum asset value is zero). Thus, a relatively smaller capital ratio is required for lognormal asset risk compared to loss risk. For the lognormal distribution, Figure 3 depicts the capital/losses as a function of the coefficient of variation of losses for the same expected policyholder deficit ratios as in Figure 1. Here, the relationship between c and k is not as linear as for the normal case, but the approximation should be reasonable for modest values of the coefficient of variation.

Figure 3
Capital/Loss vs. Expected Policyholder Deficit Ratio and
Coefficient of Variation of Losses Under Lognormal Distribution

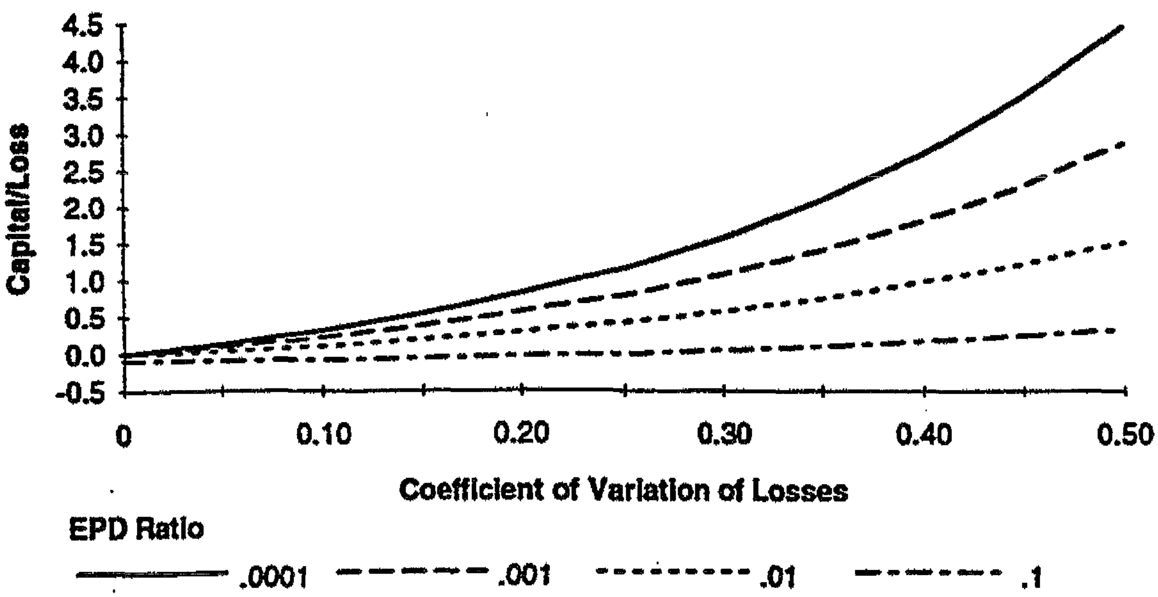
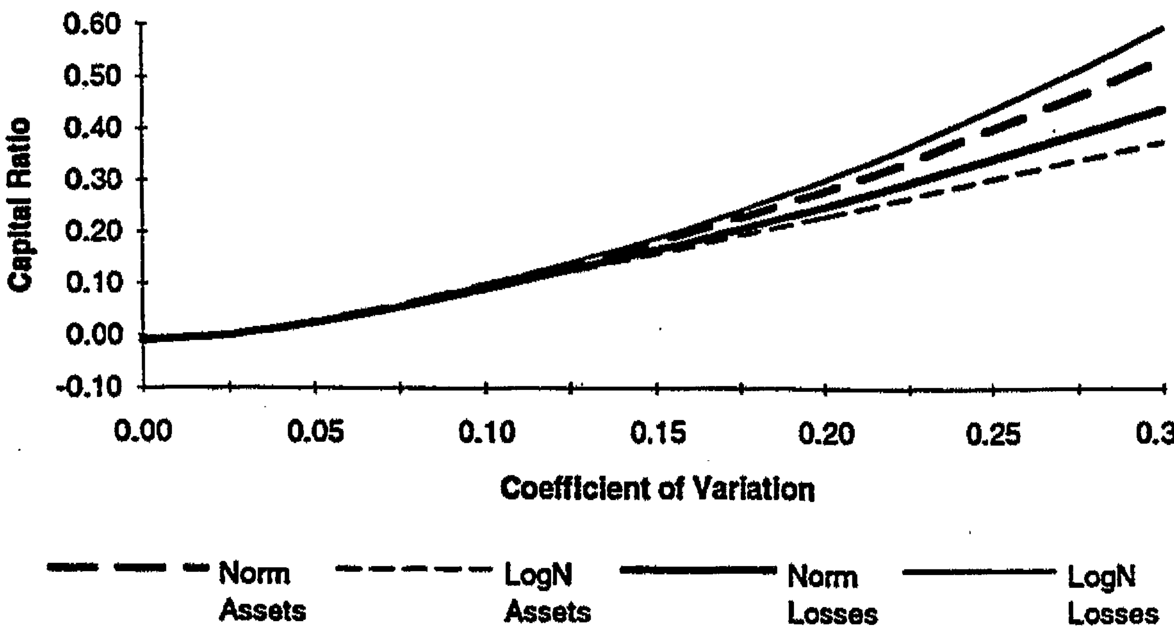


Figure 4 compares the normal and lognormal results directly for d = 0.01. Notice that the capital requirement for losses under the lognormal distribution is greater than for the normal case, with the disparity becoming greater as the coefficient of variation increases. For asset risk, the lognormal distribution requires less capital than the normal case. These relationships are due to the skewness of the lognormal model: the probability of large losses is higher than under the normal distribution and lognormal asset values cannot be negative.

Figure 4
Comparison of Asset vs. Loss Capital Ratios
for Normal and Lognormal Distributions
Expected Policyholder Deficit Ratio = 0.01



The risk-based capital methodology developed so far can be summarized thus: (1) select a particular expected policyholder deficit ratio as the solvency measure, (2) for each risk element, establish a balance sheet with an opposite riskless item (e.g., a risky asset with a riskless liability), keeping the liability amount fixed and (3) using the probability distribution of the risk element amounts, vary the assets to solve for the capital ratio (the amount of capital relative to the amount of the risk element) providing the given EPD ratio.

Risk Measurement and Time

For most risk elements, the probability distribution of realized amounts depends on the length of time until realization. Accounting standards also complicate time-dependent risk measurement.

Accounting Conventions and the Bias Problem

Valuation distortions often appear for financial statement items subject to accounting measurement prior to realization. For example, bonds and real estate will vary in market value, but their statutory or GAAP values, based on purchase price rather than current realizable value, generally remain constant until sold. Conversely, change in an accounting value per se does not connote by the accounting value) that conveys risk. To illustrate, the ultimate value of a discounted unpaid loss may be known with certainty, but although its accounting measure will change (increase) through time, there is no risk present. On the other hand, a risky unpaid loss might carry a constant reserve for several years until the uncertainty is resolved.

For solvency risk measurement, the accounting treatment should directly reveal realizable value variations. Market-value accounting, which sets all balance sheet items at current realizable value, is particularly suitable for solvency assessment, since an insurer's failure usually results in liquidation of the balance sheet or purchase of the company, both in market transactions. Using market valuation, capital is defined as the excess of the market value of assets over the market value of liabilities. Ignoring the value of intangible assets such as goodwill (which would be difficult to incorporate into a practical risk-based capital program), capital is the net liquidation (also called break-up or winding-up) value of the company.

Defining capital as the accounting book value (e.g., statutory surplus or GAAP equity) severely limits the usefulness of a risk-based capital methodology. This is due to accounting bias, which occurs when the current recorded value differs from the current realizable value.

Statutory accounting in particular allows insurers to consistently overvalue or undervalue certain financial statement elements (e.g., loss reserves are not discounted). Further, both SAP and GAAP offer inconsistent measurement between companies by allowing insurers to carry some identical items at different amounts. For example, one insurer may record its loss reserves with a margin for adverse deviation, while another may discount its loss reserves to reflect present value. Also, identical bonds purchased at different times by two insurers may be booked at different amounts.

When bias is present, an insurer's recorded capital will not represent the realizable value of assets over liabilities. For example, if reserves are carried at undiscounted value, the amount of the discount is, in effect, hidden capital. A consistent, unbiased valuation method is essential in equitably assessing insolvency risk among insurers. Consequently, when setting risk-based capital, the financial statement should be adjusted to remove bias.

Time Horizon for Risk Measurement

Under the market valuation standard, it is clear that, in order for the value of capital to change, time must elapse. For example, if the current market value of assets and liabilities are $1,200 and $1,000 respectively, the capital will have an unambiguous value of $200. No matter how risky these items are, market valuation provides a single result when the items are measured concurrently. Since the change in value of capital depends on the passage of time, it follows that insolvency risk must be measured by weighing the possible capital values at a future time. The future capital values are assets minus obligations, both balance sheet quantities. Thus, the relevant insolvency risk elements must be capable of point-in-time estimation, as opposed to a flow-through-time measurement.

In accounting theory, balance sheet items are known as stock quantities, while cash flow and income statement items are called flow quantities. Notice that the commonly-used premium-to-surplus solvency measure is the ratio of a flow to a stock quantity. For consistency in measurement, the ratio of two stock items, such as loss-related reserves and surplus, or two flow items, such as incurred losses and premiums, should be used. The capital ratios (c and c_A, for example) in this model are applied to balance sheet quantities in order to produce the required risk-based capital.

The time span between the current valuation of a financial statement item and a subsequent valuation will greatly affect the measurement of risk. For example, it is more likely for a share of stock to decline 10 percent in value in one year than in one day; similarly, liability reserves to be paid five years from now are more likely to develop adversely by 10 percent than reserves paid in the next six months. Therefore, the degree of risk depends on the time interval between valuations as well as the intrinsic volatility of the item.

The dispersion of future realizable values for many assets, notably stocks, is commonly modeled as a diffusion process.[10] Here the spread of future values becomes greater as time elapses. Similarly, it is known that the variance of unpaid liability losses increases with the time required to pay claims.

The time-dependent nature of risk is illustrated in Figure 5, where the range of possible risk element values increases with time for three periods. Here the current market value is V_0, and V_{tk} represents the kth possible value at the end of the tth period. The transition of value from one period to the next is governed by a particular probability rule. When the relationship between adjacent nodes is constant (e.g., the probability of moving from, and the rate of change from V_{11} to V_{22} is the same as from V_{23} to V_{34}), the set of possible values at any future point will have the same probability distribution, but with a regularly changing mean and variance.[11] The variance will increase through time (except for trivial cases) even though the mean could remain constant or decrease. A critical notion, therefore, is the variance per unit of time.

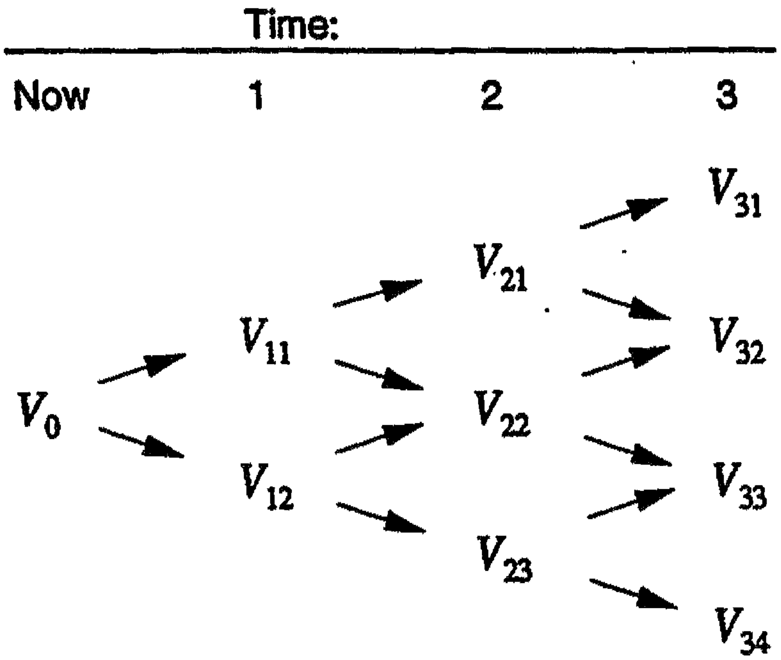
Because the variance of realizable values is time-dependent, in order to measure risk consistently, especially for different types of risk elements, it is necessary to establish a common time horizon. Extending the stock vs. reserves example, suppose that the values of the stock at the end of four years and the unpaid loss at the end of one year both have a standard deviation equal to 0.1 times their current value; both risk elements follow a normally distributed

---

[10] A diffusion process is a type of continuous stochastic process (wherein the probability structure depends on time). The prototype for diffusion processes is Brownian motion, where changes in position are independent increments. It is commonly assumed that infinitesimal stock price changes are normally distributed, producing lognormally distributed stock returns (see Brockett and Witt, 1990, and Cummins, 1988, for additional details).

[11] For example, suppose V_{tk} is a particular node on the lattice and the two subsequent possible values are $V_{tk} + a$ and $V_{tk} - a$ (for all t and k, where a is a constant) with probabilities p and $1 - p$. Then the mean will increase by the constant amount $a(2p - 1)$ and the variance by the constant amount $4p(1 - p)a^2$ over each period. If the two subsequent values are *multiples* of V_{tk}, then the mean and the *second moment* each increase by a constant factor each period.

Figure 5
Time Dependence of Risk



diffusion process. Thus, the standard deviation is proportional to the square root of the elapsed time and the reserve standard deviation is twice the stock standard deviation at any particular time. Table 4 illustrates the results for \$1,000 of each item with capital equal to \$100. The two items do not have the same risk: since the reserve standard deviation is higher than that of the stock for a common time horizon, there is a pronounced difference in both the ruin probability and expected insolvency cost.

Table 4
Two Normally Distributed Risk Elements with Different Risk per Time Unit—Beginning Capital = 100

	Value	*Standard Deviation*		*Probability of Ruin (10 Percent Adverse Change)*		*Expected Policyholder Deficit*	
Time	Now	1 Year	4 Years	1 Year	4 Years	1 Year	4 Years
Stock	1,000	50	100	0.023	0.159	0.47	9.26
Unpaid Loss	1,000	100	200	0.159	0.309	8.33	39.56

Note: Equations (4a) and (4b) are used to calculate the expected policyholder deficit values. The capital ratios are 0.10, and the coefficients of variation are equal to (standard deviation)/1,000. Ruin probabilities are determined from the cumulative normal distribution evaluated at [–100/(standard deviation)]. The four-year stock and one-year unpaid loss standard deviations are assumed to be 10 percent of the initial \$1,000 value of each risk element.

Capital Adjustment over a Multiperiod Time Horizon

The parties to the insurance contract are concerned about insolvency of the insurer over a very long time horizon. However, by defining the solvency criterion as the expected policyholder deficit ratio per fixed unit of time, a sufficient and consistent solvency protection can be achieved simply by meeting the expected policyholder deficit criterion continually over a short time horizon, equal to the interval between risk-based capital evaluations.

A numerical example will demonstrate this concept. Assume that an insurer has an initial unbiased loss reserve of $1,000. The actual loss payment will occur in three years. Initial assets are $1,100 and remain fixed until the loss is paid; to simplify the example, interest is zero. During each year, information is gathered that enables us to reevaluate our expectation of ultimate loss. The nature of this information is limited, such that the reserve can change, with equal probability, each year up or down by 20 percent of the previous reserve. Thus, the reserve sequence through time is a simple binomial stochastic process.[12] Table 5 shows the progression of possible reserve values, along with the associated probabilities. The EPD when the loss is paid is $98.

Table 5
Time-Dependent Solvency Measurement for
Unpaid Losses Using a Binomial Stochastic Process

Time (t)	*Now (0)*	*1*	*2*	*3*
				1,728 (0.125)
			1,440 (0.25) <	
		1,200 (0.5) <		**1,152 (0.375)**
	1,000 <		960 (0.50) <	
		800 (0.5) <		768 (0.375)
			640 (0.25) <	
				512 (0.125)
Expected Loss	1,000	1,000	1,000	1,000
Assets	1,100	1,100	1,100	1,100
EPD if Liquidated		50	85	98
EPD Ratio		0.050	0.085	0.098

Note: The above values represent either reserve or paid loss amounts, depending on the duration of ultimate loss; the probability of each reserve/payment value is shown in parentheses. The expected policyholder deficit (EPD) is measured now (time 0) for payment at time t. The reserve values in bold represent a policyholder deficit for a three-year horizon if the insurer is liquidated or the claim is paid at that point.

Now assume that the loss follows the same stochastic process but is paid in two years. In other words, the possible paid loss amounts at Year 2 equal the reserve amounts at Year 2 for the three-year case. The two-year EPD is $85 (Table 5). Similarly, under the same stochastic process, a loss paid in one year has a $50 EPD, or a 5 percent EPD ratio.

---

[12] By using sufficiently small time intervals, a simple binomial structure can replicate a continuous diffusion process. In particular, if the relationship between adjacent nodes is multiplicative (as in the example here), the result will tend toward a lognormal distribution. Cox, Ross, and Rubinstein (1979) and Trigeorgis (1991) apply the binomial method to evaluate time-dependent contingencies. Some of the numerical results for normal and lognormal distributions in this article were obtained using this technique.

With the amount of assets fixed until settlement, the expected insolvency cost increases with the length of time required to pay the loss, since the spread of possible loss values widens. However, under a typical risk-based capital program, an insurer would be required to adjust its capital (and thus its assets) annually to maintain the minimum requirement. This adjustment process transforms the multiperiod problem into a single-period one.

The annual adjustment procedure is equivalent to liquidating the insurer at the end of each period. When this occurs, the policyholder deficit equals the loss reserve minus the assets. This is true despite the fact that the ultimate loss payment will be unknown at that time. For instance, suppose at Year 1 the reserve is $1,200 and assets are $1,100. The insurer is liquidated, releasing its assets to the policyholders. The ultimate losses will be 1,728, 1,152, and 768 with respective probabilities of 0.25, 0.50, and 0.25, yielding respective deficits of 628, 52, and –332. The –332 deficit represents a profit, because the policyholders have, in effect, self-insured for a $1,100 premium.[13] The expected deficit is $100, the same as if the $1,200 reserve were a payment.

Because in this example the probability distribution for the annual percentage reserve change is independent of the elapsed time, a 5 percent one-year forward EPD ratio is achieved when capital is 10 percent of the current expected loss. So, assuming an $800 reserve outcome at Year 1, the capital will increase to $300. We can shed most of it, since only $80 of capital is now needed. As shown in Table 6, the expected deficit over the next period (Year 1 to Year 2) is 0.5(960 – 880) + 0.5(0) = $40, or 5 percent of the $800 expected loss.

The capital adjustment can be repeated each period until the loss is finally paid. Thus, although the ultimate expected deficit depends on the duration of the unpaid loss as well as its volatility per unit of time, by measuring the expected policyholder deficit over a fixed time span, we can eliminate the effect of different risk element durations. Notice that without a common time horizon, additional assumptions concerning insurer behavior are required. For example, the holding period for asset sales and capital management strategy must be specified.[14]

In the above simple example, the reserve at any point is equal to the expected paid loss. In a more realistic situation, the market value of the reserve will depend on prevailing interest rates as well as the risk of adverse development. The market value can be obtained by valuing the unpaid losses as in a

[13] To avoid having a profit, the policyholders could reinsure 11/12 of the loss for the $1,100 premium. In this case the reinsurer's expected loss equals the premium, so the deal is fair. The policyholder deficit remains at $100.

[14] In the example of Table 5, suppose that the insurer always adjusts capital so that the one-year EPD ratio is 5 percent and defaults only if the loss paid in Year 3 exceeds the Year 2 assets. In this case, the ultimate EPD will be $50, the same as the EPD over just the first year. But if the insurer always defaults when the reserve or loss payment exceeds the assets and withdraws capital when the reserve becomes smaller (still maintaining the 5 percent one-year EPD ratio), then the ultimate EPD will increase to $78.

commutation with the policyholder or a loss reserve portfolio transfer to another insurer or reinsurer.[15]

Table 6
Time-Dependent Solvency Measurement for Unpaid Losses
Using a Binomial Stochastic Process Viewed One Year Later
Given that Expected Unpaid Loss is $800

Time (t)	*Now (1)*		*2*			*3*	
						1,152	**(0.25)**
			960	(0.5)	<		
	800	<				768	(0.50)
			640	(0.5)	<		
						512	(0.25)
Expected Loss	800		800			800	
Assets	880		880			880	
EPD if Liquidated			40			58	
EPD Ratio			0.050			0.085	

Note: The format is the same as Table 5's. The present time is one year later (time 1). The expected policyholder deficit (EPD) is measured at present for a liquidation at time t. The reserve values in bold represent a policyholder deficit if the insurer is liquidated at that point (or at the third year, when the actual claim must be paid). Capital has been reduced to provide a one-year forward EPD ratio of 5 percent.

The capital adjustment process can be described symbolically by extending the earlier notation. Let $\tilde{A}_t$, $\tilde{L}_t$, and $\tilde{C}_t$ be random variables denoting respective assets, liabilities, and capital at time t. The amounts of these variables are known at present: A_0, L_0, and C_0, with $A_0 = (1+c)L_0$. Further, let $\tilde{A}_1 = (1+\tilde{r})A_0$ and $\tilde{L}_1 = (1+\tilde{g})L_0$, where $\tilde{r}$ and $\tilde{g}$ are random variables denoting the annual return on assets and the annual rate of change in value of the liabilities (i.e., the expected value of $\tilde{g}$ is a risk-adjusted discount rate).

An important variable is $\tilde{C}_1$, the amount of capital at the end of one period. Define $\tilde{c}_1 = \tilde{C}_1/L_0$ as the amount of capital relative to the original expected loss. Then $\tilde{c}_1 = c + (1+c)\tilde{r} - \tilde{g}$, and we have the one-period expected policyholder deficit ratio

$$d_1 = \int_{-\infty}^{0} -zp(z)dz, \tag{6}$$

where $p(\bullet)$ = the density of $\tilde{c}_1$. For a given value of d_1, then, we solve the formula for the beginning capital/loss ratio c. The Appendix applies this formula to derive EPD ratios for the normal distribution when assets and liabilities are

[15] The market value of loss reserves would also implicitly assume that there is no insurer default; otherwise, it would depend on the capital levels of individual insurers and thus could not be a market value. In the absence of a ready market for the trading of loss reserves, market values can be approximated by discounting using a risk-adjusted interest rate (see Butsic, 1988, and D'Arcy, 1988, for details).

both random variables (the negative sign in the integral converts negative capital amounts to positive policyholder deficits).

To summarize, the capital adjustment process guarantees that the policyholders will maintain or exceed the same expected policyholder deficit ratio over each successive year. Therefore, by choosing a common time horizon, a consistent level of policyholder safety can be provided without regard to the actual duration of the risk element. The key requirements are that we know the time-dependent nature of the future realizable values, that the insurer can liquidate its assets and liabilities at each evaluation point (although it does not have to do so), and that insurers will add capital when needed. Notice that it is not necessary for the time-based probability structure to be uniform, as in the preceding example—we only need to know the current market value and the probability distribution of future market values at the *next* valuation date.[16]

The Insurer as a Going Concern

The preceding discussion treated the runoff of an insurer; in other words, it did not consider the risk from policies (both new and renewal) becoming effective in the future. This contingency is considered a major risk element, with rapid growth in business a primary cause of property-liability insolvencies. (A study by A. M. Best, 1991, showed that over the period 1969 through 1990, 21 percent of property-liability insolvencies were caused by rapid premium growth.) Nevertheless, the periodic capital calibration to assure a minimum expected policyholder deficit ratio will also work for an insurer as a going concern, continually writing new business.

To incorporate the future business into the calibration procedure, we still assume the capability of liquidating the insurer at the end of each successive period. However, now the year-end capital $\tilde{C}_1$ will be affected by both the runoff of the initial balance sheet (assets and liabilities A_0, L_0) and by the period-ending value of the business added during the interval. Let P be the premium (net of expenses), assumed to be entirely collected just after the beginning of the period, and let $\tilde{L}_P$ be the loss from the added premium, assumed to be incurred at the end of the period and paid at the end of a subsequent period.[17] Then the end-of-period assets and liabilities are $\tilde{A}_1 = (A_0 + P)(1+\tilde{r})$ and $\tilde{L}_1 = L_0(1+\tilde{g}) + \tilde{L}_P$, where $\tilde{r}$ and $\tilde{g}$ are as previously defined (we assume that the premiums are invested in the same assets as A_0; also the premium is known in advance and thus is not a random variable).[18]

---

[16] The current market value will embody the market's knowledge of all possible future values. Thus, knowing (or being able to estimate) the market value will, in effect, provide information concerning the forward probabilities.

[17] The beginning and end-of period assumptions are not necessary, but help to clarify the analysis. A more refined approach would place the premiums and incurred losses in the middle of the period.

[18] This assumption is reasonable, since we can predict the amount of premium for the next year fairly well from the insurer's business plans and knowledge of the current market. However, it would be increasingly difficult to accurately forecast premiums longer into the future. Even

By choosing a constant p we relate the premium to the initial capital: $P = pC_0 = pcL_0$. Also, define a random variable $\tilde{b}$ equal to the incurred loss ratio: $\tilde{L}_P = \tilde{b}P = \tilde{b}pcL_0$. Using the previous definition of the period-ending capital to the initial liabilities, we get

$$\tilde{c}_1 = c(1+p) + [1+c(1+p)]\tilde{r} - pc\tilde{b} - \tilde{g}. \tag{7}$$

Finally, we solve equation (7) for the value of c needed to determine d_t.

Equation (7) is a linear function of three important random variables comprising the bulk of the risk facing the typical property-liability insurer. We have already discussed the role of the asset and loss reserve risk, represented by $\tilde{r}$ and $\tilde{g}$. These can be modeled as diffusion processes. The incurred losses (represented by $\tilde{b}$), as well as a portion of the unearned premium reserve, have components that are qualitatively different from the other risks. In particular, property coverages are subject to catastrophes, which are highly unpredictable, and, being paid quickly, cannot be modeled as diffusion processes (Cummins, 1988, addresses catastrophes by modeling them as a jump process added to a diffusion process). I will return to the three risk categories below in a discussion of correlation of risk elements.

As stated earlier, risk elements are balance sheet quantities. Although the above premiums and incurred losses are income statement items occurring between evaluations, their present value is a balance sheet quantity. A true market valuation would include the present values of future premiums collected and losses and expenses paid arising from business not yet written. But since assessing the worth of an insurer's future business would be a formidable task for a regulator using only public financial statements (except for business added in the upcoming year), this item will be ignored as a risk element for practical risk-based capital applications.

Because the present value of future business is a balance sheet quantity, the capital adjustment process guarantees that the policyholders will maintain at least the minimum EPD ratio each year, even if more exposures occur between evaluation points.

Insolvency Cost as a Financial Option

The model for the expected policyholder deficit ratio as developed so far is nearly complete. However, the *present value* of the policyholder deficit also must be considered: one dollar of forfeited claim a year from now is worth less than one dollar lost now. Since we are evaluating the expected policyholder deficit occurring one year hence, its value is reduced by $1/(1+i)$, where i is a default-free interest rate with a one-year duration (i.e., a one-year principal-only treasury note). For the remainder of this article, it is assumed that the EPD is measured at present value.

---

though there is considerable risk to the long-range premium forecast, the insurer does not need extra capital *now* to offset the future uncertainty. The annual risk-based capital calibration process

With the addition of the present value concept, the expected policyholder deficit is now completely analogous to a *financial option* having a one-year maturity. Return to the example with a 50/50 chance of a \$1,200 or \$800 loss reserve value at the end of one year, with \$1,100 in assets at the end of the year. The EPD valued at that point is \$50. With 8 percent interest, the present value of the EPD is \$50/1.08 = \$46.30.

Now suppose that a share of stock has a current price of \$1,000 but will be worth either \$1,200 or \$800 in one year with equal probability. An option to buy (a call option) one share a year from now for \$1,100 (the exercise price) is available.[19] If the stock turns out to be worth \$1,200, the option is worth \$100. If its price is \$800, the option would not be exercised (if so, the holder would lose \$300) and thus its value would be zero. The expected option value at the exercise date is therefore \$50 = 0.5(100) + 0.5(0). Its present value at the same 8 percent interest rate is \$46.30, which is identical to the value of the expected policyholder deficit.

Thus, for a liability risk element paired with riskless assets, the value of the EPD is equivalent to that of a call option on the losses with an exercise price equal to the value of the assets (A_1) at the end of the year (see Table 7). In effect, because liabilities may exceed the insurer's assets, its policyholders have given the insurer's owners the *option* to abandon full payment of claims. The legal concept of corporate limited liability (nonassessment for mutual policyholders/owners) creates this option. Garven (1992) discusses the pricing and incentive implications of the "limited liability" option for stock and mutual insurers.

Table 7
Equivalence Between Call Option on Stock and Policyholder Deficit—Risky Liabilities and Riskless Assets

Stock		*Insurance*
S_0: Current Stock Price	↔	L_0: Current Liability Value
S_1: Stock Price in One Year	↔	L_1: Liability Value in One Year
E_1: Exercise Price	↔	A_1: Asset Value in One Year
E_0: Present Value of Exercise Price	↔	A_0: Current Asset Value
$E_0 - S_0$	↔	C_0: Current Capital Value
$\text{Max}[0, S_1 - E_1]$: Option Value When Exercised	↔	$\text{Max}[0, L_1 - A_1]$: Policyholder Deficit

For an asset risk element (paired with a riskless liability), the EPD is equivalent to the value of a put option on the ending assets implicitly given by the

[19] An option exercisable only at the expiration date, as we have assumed here, is called a *European* option. An option exercisable at any time until the expiration date is called an *American* option. Technically, policyholders have implicitly written American options against their claims, since the insurer's owners can "exercise" during the year, rather than at the evaluation dates. However, the difference in option value (present value of the EPD) would not be significant.

policyholders (see Table 8). Here, if the asset value (stock price) in one year is less than the liability value (exercise price) in one year, the difference is *put* to the policyholders (the option seller). A put option equivalence for both a risky asset and a risky liability is shown in Table 9 (see Brealy and Myers, 1988, and Cox and Rubinstein, 1985, for discussions of option relationships).

Table 8
Equivalence Between Put Option on Stock and Policyholder Deficit—Risky Assets and Riskless Liabilities

Stock		*Insurance*
S_0: Current Stock Price	$\leftrightarrow$	A_0: Current Asset Value
S_1: Stock Price in One Year	$\leftrightarrow$	A_1: Asset Value in One Year
E_1: Exercise Price	$\leftrightarrow$	L_1: Liability Value in One Year
E_0: Present Value of Exercise Price	$\leftrightarrow$	L_0: Current Liability Value
$S_0 - E_0$	$\leftrightarrow$	C_0: Current Capital Value
$\text{Max}[0, E_1 - S_1]$: Option Value When Exercised	$\leftrightarrow$	$\text{Max}[0, L_1 - A_1]$: Policyholder Deficit

Table 9
Equivalence Between Put Option on Stock and Policyholder Deficit—Risky Assets and Risky Liabilities

Stock		*Insurance*
S_0: Current Stock Price	$\leftrightarrow$	C_0: Current Capital Value
S_1: Stock Price in One Year	$\leftrightarrow$	C_1: Capital Value in One Year
E_1: Exercise Price	$\leftrightarrow$	Zero
E_0: Present Value of Exercise Price	$\leftrightarrow$	Zero
$S_0 - E_0$	$\leftrightarrow$	C_0: Current Capital Value
$\text{Max}[0, E_1 - S_1]$: Option Value When Exercised	$\leftrightarrow$	$\text{Max}[0, -C_1]$: Policyholder Deficit

The idea of insurer solvency cost being a financial option is a fairly recent development, trailing the rapid growth of stock option trading in the 1970s. For a more thorough treatment of the topic, see Doherty and Garven (1986), Cummins (1988), Derrig (1989), and Garven (1992). In particular, Cummins shows that the value of the risky asset-liability put option (here, the EPD) is the fair risk-based guaranty fund premium. The Appendix derives EPD ratios for the lognormal distribution by using call and put option equivalents to the respective EPDs for risky losses and assets.

We now have a complete capital-setting model for individual risk elements: determine how much capital per unit of risk element satisfies a standard value of the one-year discounted expected policyholder deficit ratio.[20] For a liability

[20] Due to the annual horizon of a practical risk-based capital program, taking the present value of the EPD will not change the *relative* capital ratios needed for one risk element versus another. This is because the same riskless interest rate should be used for all risk elements. Thus, for example, if the EPD ratio standard is set by requiring a specified percentile of insurers failing to reach the standard, then taking the present value of the EPD is not necessary.

risk element, we assume that the related asset is riskless, with annual return $\tilde{r} = i$. In parallel fashion, a risky asset is paired with a riskless liability, whose market value grows at an annual return of $\tilde{g} = i$. The next section extends the results to the more likely case where both assets and liabilities are risky.

Correlation and Independence of Risk Elements

The preceding sections have demonstrated how risk-based capital for each risk element can be calculated separately by treating each element in the balance sheet of a mini-insurer. As shown below, in order to combine the risk capital for the separate elements, one cannot simply add their required capital amounts together unless the risk elements are perfectly correlated (with the proper sign).

A Numerical Illustration

Suppose that we have a line of business with riskless assets and risky losses, which can have only two possible realizable values. The values and their probabilities are given in Table 10. The desired expected policyholder deficit ratio is 1 percent. The risk-based capital needed for this degree of protection is easily calculated at $2,900.

Assume that another line of business has an identical loss distribution, directly correlated with the first: if a $2,000 loss occurs for the first line, the

Table 10
Insurer with Two Independent Lines of Business with
Same Unpaid Loss Distribution—Asset Amount Is Certain

	Asset Amount	*Loss Amount*	*Capital Amount*	*Probability*	*Claim Payment*	*Deficit*
Single Line						
	6,900	2,000		0.6	2,000	0
	6,900	7,000		0.4	6,900	100
Expected Value	6,900	4,000	2,900		3,960	40
Capital/Loss		0.725				
EPD Ratio		0.010				
Two Independent Lines						
	13,800	4,000		0.36	4,000	0
	13,800	9,000		0.48	9,000	0
	13,800	14,000		0.16	13,800	200
Expected Value	13,800	8,000	5,800		7,968	32
Capital/Loss		0.725				
EPD Ratio		0.004				

Note: When capital for the two-line case is reduced to 5,500, the expected policyholder deficit (EPD) ratio becomes 0.16(14,000 − 13,500)/8,000 = 0.010; the capital/loss drops to 0.687.

same amount occurs for the second line; similarly, a $7,000 amount will occur concurrently for both lines. The effect of combining the two lines is the same as if we now had a single line twice as large as the original single line. The amount of capital per unit of expected loss needed to provide the 1 percent EPD ratio remains the same at 0.725.

Now suppose that the two lines are statistically *independent:* the value of the loss for one line does not depend on the value for the other. Combining the two lines and adding their separate $2,900 risk-based capital amounts creates the EPD for the composite line (see Table 10). Here the probability of a loss exceeding assets is reduced from 0.40 to 0.16, and the $32 expected deficit for the combined lines is less than the sum of the individual expected deficits ($80). This produces a 0.004 protection level, compared to the 0.010 value for the separate lines. To reach the same 1 percent level as before, we need *less* capital than obtained by adding the separate amounts of risk-based capital: only $5,500 of capital is required, which is $480 less than the $5,980 needed when the losses are correlated. The capital ratio to loss drops from 0.725 to 0.687.

For a discrete loss distribution, with assets certain, the EPD for the sum of n independent equal losses is given by $D_L = \sum_{x>A} p_n^*(x)(x-A)$, where $p_n^*(\bullet)$ is the n-fold convolution of the probability density for the losses $(0 \leq x < \infty)$.

The reason for the reduced capital requirement through independence of risk elements is the *law of large numbers*. When losses are independent of each other, a small line of business will need a relatively large amount of capital per unit of loss, while a larger one requires a much smaller capital ratio. Practically, however, there is a limit to the risk reduction allowed by the law of large numbers. The mean or other parameters of the loss distribution are rarely known with certainty, introducing systematic or parameter risk affecting all exposures. Thus, even an insurer with a very large homogeneous book of business will remain subject to considerable uncertainty, and consequent capital needs.

Correlation Under the Normal Distribution

Although the preceding numerical example illustrates the capital reduction due to independence of risk elements, one must be careful not to generalize regarding the degree of reduction. For example, using a 0.1 EPD ratio, the capital requirement drops to $2,000 for the single line of business. The *combined* capital need drops to $1,000 for the two independent lines—less capital than for a single line. This effect is due to using a discrete probability distribution with a limited range of outcomes. More robust conclusions can be reached by analyzing a continuous probability model, such as the normal distribution.

The normal distribution has the important property that sums of normal random variables are themselves normal random variables with additive means and easily-computed variances. Table 11 provides the mean and variance for

Table 11
Mean and Variance of Capital with Two Normally Distributed Assets or Liabilities

Random Variables	*Mean*	*Variance*
Two Assets	$C = A_1 + A_2 - L$	$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$
Two Liabilities	$C = A - L_1 - L_2$	$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$
Asset and Liability	$C = A - L$	$\sigma^2 = \sigma_A^2 + \sigma_L^2 - 2\rho\sigma_A\sigma_L$

Note: σ_1 and σ_2 denote the standard deviations of risk elements 1 and 2 (either assets or liabilities). For the asset and liability combination, σ_A is the total asset standard deviation, and σ_L is the total liability standard deviation. The correlation coefficient between risk elements is ρ.

the composite of two normally distributed assets ($\tilde{A}_1$ and $\tilde{A}_2$), or two liabilities ($\tilde{L}_1$ and $\tilde{L}_2$), or an asset and a liability ($\tilde{A}$ and $\tilde{L}$).

With perfect positive correlation ($\rho = 1$), $\sigma = \sigma_1 + \sigma_2$ for risk elements on the same side of the balance sheet or $\sigma = \sigma_A - \sigma_L$ for assets and liabilities. With perfect negative correlation ($\rho = -1$), $\sigma = \sigma_1 - \sigma_2$ or $\sigma = \sigma_A + \sigma_L$. When the elements are independent, $\rho = 0$, and thus $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ and $\sigma = \sqrt{\sigma_A^2 + \sigma_L^2}$ for the two cases. As shown in the Appendix, the formula for the EPD ratio with normally distributed combined risk elements is identical to that for individual elements, equations (4a) and (4b).

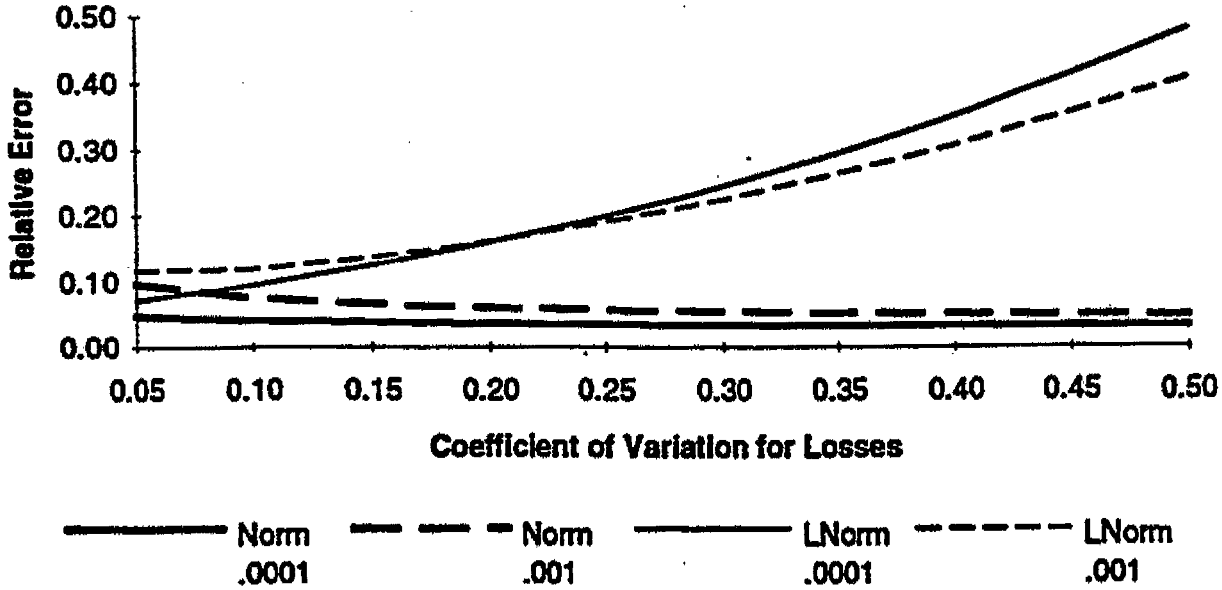
Here c is the capital to loss, k is the total standard deviation divided by the total expected loss L, and D is the total expected policyholder deficit. Similarly, the approximate lognormal (assuming that the sum of two lognormal variables is lognormal) EPD ratio for combined risk elements is identical to equations (5a) and (5b).

As indicated above, for the normal and lognormal distributions, the relationship between c and k is approximately linear for a fixed EPD ratio d and small k. Since $c = -d$ when $k = 0$ (no risk), we have $c \cong ak-d$ for some constant a. Assuming that a high level of protection is desired (d less than 1 percent or so), we further simplify the relationship to $c \cong ak$.

Since the total capital C equals cL, and the total standard deviation σ equals kL, it follows that if $c = ak$, then $C = akL = a\sigma$. Therefore, the risk-based capital for the total of separate risk elements is proportional to their combined standard deviation. Risk capital for perfectly correlated items can be added (or subtracted, depending on whether the correlation is positive or negative or whether the items are on the same side of the balance sheet). Risk capital for independent (and partially correlated items) can be combined according to the square root of the sum of the squares of their standard deviations, plus twice the product of their standard deviations and the correlation coefficient. We will refer to this as the *square root rule*.

The relative error in using the square root rule for two independent loss risk elements of the same size and standard deviation is shown in Figure 6. Here the relative error is defined as (approximated minus true capital)/(true capital). For the normal distribution, the error decreases as the EPD ratio decreases and as the risk increases. For a reasonable (i.e., 0.001) protection level, the error is less than 10 percent. For the lognormal model, the error increases with increas-

Figure 6
Relative Error Using Square Root Approximation
for Losses Under Normal and Lognormal Distributions
Expected Policyholder Deficit Ratios = 0.0001 and 0.001



ing risk and tends to be much larger than with the normal distribution. However, for modest values of k, the linear approximation may be acceptable. Notice that for both distributions the relative error is positive, meaning that the square root rule overstates required capital.[21] Thus, the approximation is conservative.

To illustrate, suppose that we have two independent normally-distributed lines of business, each with a $1,000 expected loss and $200 standard deviation. For a 0.001 expected policyholder deficit ratio, each requires $438 of capital in isolation (from equation (4a) with k = 0.2 and c = 0.438). When the lines are combined, the overall standard deviation increases to $283, but the capital ratio declines to 0.292, giving $584 in capital when applied to the $2,000 total expected losses. The square root rule produces $620 = $438\sqrt{2}$, which is about 6 percent more than the exact calculation yields.[22]

A parallel calculation with the lognormal distribution, using equation (5a) shows a 16 percent error: the true required capital is $700, compared to $812 from the square root rule.

The square root rule can be extended to incorporate more than two risk elements. The total capital C is a function of the individual element risk capital amounts C_i and the separate correlation coefficients between each pair of n risk

---

[21] The reason for this is apparent from Figure 3, which shows the capital ratio c to be a *concave upward* function of k. In other words, $d^2c/dk^2 > 0$. Adding an equal, independent loss reduces k to $k' = k/\sqrt{2}$, but the capital ratio drops by even more, to $c' < c/\sqrt{2}$.

[22] Because the error in using the square root for the normal and lognormal distributions overstates the combined amount of capital needed, a closer fit could be had by using a root higher than two. For instance, in the normal example given, using a

elements (notice that the sign of the correlation coefficient depends on which side of the balance sheet the two items reside):

$$C = \left[\sum_{i=1}^{n} C_i^2 + \sum_{i \neq j}^{n} \rho_{ij} C_i C_j \right]^{\frac{1}{2}}. \tag{8}$$

Practical Application of Correlated and Independent Risk Elements

The preceding analysis has shown the effect of correlation between risk elements. Some examples of balance sheet items having varying degrees of correlation are presented in Table 12. In general, reinsurance transactions create a high degree of correlation between ceding and assuming parties. Ownership of insurance subsidiaries (affiliates) or stock also produces highly correlated values. Where it is difficult to determine the numerical correlation between items, a practical approach would be to judgmentally peg the correlation at zero, 1, or −1, whichever is closest to the perceived value.

Table 12
Independent and Correlated Risk Elements

Correlation	*Asset/Asset*	*Liability/Liability*	*Asset/Liability*
Positive	Common Stock/ Preferred Stock, Common Stock/Bonds	Loss Reserve/LAE Reserve	Bonds/Loss Reserve
Zero	Cash/Real Estate	Liability Loss Reserve/ Property Unearned Premium Reserve	Common Stock/ Unearned Premium Reserve
Negative	Common Stock/ Put Options	Loss Reserve/Income Tax Liability, Loss Reserve/ Dividend Reserve	Property-Liability Stock/Loss Reserve, Reinsurance Recoverable/ Loss Reserve

We can demonstrate the effect of independent and correlated risk elements by constructing a numerical example. Table 13 presents risk elements from a hypothetical insurer's balance sheet at market values. The capital ratios assume a 0.005 EPD ratio and are based roughly on empirical data.

All risk elements are assumed to be lognormally distributed, and the EPDs are discounted at an 8 percent riskless interest rate (equations (5a) and (5b), adjusted for the interest rate, are used for this calculation). The loss reserve, equal to the present value of the expected payments, also includes the loss expenses and the liability portion of losses arising from the unearned premiums. The affiliate stock risk is assumed to be the same as for noninsurance stock.

The 20 percent stock capital factor arises from using a 0.168 standard deviation of 1946 to 1989 annual returns from Ibbotson and Associates (1990). Based on the same source, we have used a 0.06 annual standard deviation for bonds (the corporate bond standard deviation is 0.098 for a 20-year maturity;

Table 13
Risk-Based Capital (RBC) Calculation Using the Square Root Rule: Input Assumptions

Risk Element	*Amount*	*Capital Ratio*	*RBC*
Stocks	200	0.20	40
Bonds	1,000	0.05	50
Affiliates	100	0.20	20
Loss Reserve	800	0.40	320
Property UPR	100	0.20	20
Total			450

Correlated Risk Elements		*Correlation Coefficient*
Stocks	Bonds	0.2
Stocks	Affiliates	1.0
Bonds	Affiliates	0.2
Bonds	Loss Reserve	0.3
Affiliates	Loss Reserve	-1.0

adjusting for a more typical property-liability insurer's duration gives a lower value), producing an approximate 5 percent capital ratio.

The loss reserve capital ratio is based on a study of loss ratio variation by Derrig (1989), who used a sample of workers' compensation and private passenger auto loss ratios from 51 insurers over the period 1976 through 1985 (since calendar-year losses were used, the variance should be similar to that for loss reserves). The combined annual variance was 0.059, which was judgmentally reduced to 0.045, reflecting a greater variance in the unpaid loss tail; the variance is lowered when the loss is brought to present value. This produces a capital ratio (to the discounted loss) of about 0.40.

Notice that a further adjustment would be needed to convert the loss capital factor for application to an *undiscounted* loss reserve: using a 16 percent reserve discount (three-year loss duration at a 5 percent risk-adjusted interest rate), the required statutory surplus is $(1 + 0.40)(1 - 0.16) - 1 = 0.176$ times the undiscounted reserve. This illustrates the bias problem inherent in statutory accounting—if in this example the discount is greater than 29 percent, then the required capital is negative! Market-value accounting would avoid this problem.

The sum of the separate risk-based capital amounts is $450. This value assumes that *all* items are fully correlated, ignoring any independence or partial covariance between the items. The pairs of elements that are assumed to be correlated for this example are presented in Table 13.[23]

---

[23] The correlation coefficient for common stocks and bonds is based on the Ibbotson and Associates (1990) data. The other correlation coefficients are determined judgmentally. In practice, due to limited data, for a particular insurer one would have to rely heavily on industry results (perhaps modified with judgment) to achieve accurate correlation estimates.

The property unearned premium reserve is independent of all other items. Notice that the bonds/reserve correlation coefficient is positive due to the parallel change in value from interest rate movements; because these two items are on opposite sides of the balance sheet, their joint movement will reduce total risk.[24] Similarly, the negative sign of the affiliates/reserve correlation coefficient indicates that these opposing items will increase total risk when combined.

Applying equation (8), the sum of the squares of the separate risk capital amounts is 107,300. The sum of the cross products (each of the above pairs appears twice) of the capital amounts times their correlation coefficients equals 6,000. Thus, the approximate total risk capital is $\$337 = \sqrt{113{,}300}$. If all the risk elements were independent, the total required capital would be only $\$328 = \sqrt{107{,}300}$.

The impact of the bond/reserves covariance can be found by setting the correlation coefficient to zero: here the total risk capital increases to $351. Thus, the effect of their correlation is to reduce required capital by $14.

Notice that, for this example, if the correlations between risk elements are not known, the potential error associated with assuming that they are independent when they are fully correlated is about 27 percent of the true risk-based capital. Similarly, the error from assuming that they are fully correlated when they are independent is about 37 percent. This degree of uncertainty can have a greater impact on the insurer's total RBC than the uncertainty in assessing the RBC for a particular risk element. For example, if the risk elements are fully correlated, a 100 percent error in specifying the bond RBC (i.e., $0 or $100 instead of the indicated $50) creates only an 11 percent error in the total risk-based capital. Thus, knowing the degree of correlation between risk elements can be as important as knowing the risk of individual items.

A more sophisticated risk-based capital calculation would divide the risk elements into additional categories and might include a provision for the value of future business.

Summary and Conclusion

This article has addressed insolvency risk measurement for a risk-based capital program, with the following key results:

1. The relevant measure of solvency is the present value of the expected policyholder deficit as a ratio to the expected loss. This value is equivalent to a put option held by the insurer's owners and equals a fair risk-based guaranty fund premium. By requiring sufficient capital to meet or exceed a common

[24] The correlation methodology provides a means of allowing for matching of asset and liability durations. If the durations of fixed maturity assets and loss payments were equal, and the movements in value were due solely to interest rate fluctuations, then a (negative) 100 percent correlation coefficient would be appropriate.

expected policyholder deficit ratio standard for each insurer, policyholders are assured a consistent minimum level of protection.

2. To remove measurement bias caused by accounting conventions and varying insurer practices, the valuation standard for risk-based capital application should determine a market value for each risk element.

3. The major components of insurance risk are time-dependent: the longer the time to realization, the greater the risk. This relationship is particularly important for stocks, bonds, loss reserves, and loss adjustment expense reserves. In order to properly compare risk between these items, a common time horizon must be used.

4. The expected policyholder deficit ratio is based on expected market values at the end of each risk-based capital valuation interval (generally one year). When risk capital levels can be set periodically, with sufficient time for insurers to add capital where necessary, there is no need for additional capital to absorb fluctuations in value beyond the valuation interval. Capital is not required now for distant contingencies.

5. The risk-based capital for an insurer will always be less than the sum of the separate RBC amounts for each risk element, to the extent that all the elements are not fully correlated. By assuming a normal or lognormal distribution, an approximate method for combining risk capital is the square root of the squared individual risk-based capital amounts plus additional terms involving the correlation coefficients. Knowing the degree of correlation between risk elements can be as important as knowing the risk of individual items.

Although this article has viewed the solvency problem from a regulator's perspective, the concepts here could readily be applied to an insurer's in-house capital management. For example, the insurer might want a consistent level of capital higher than the regulatory target RBC. Or, in the absence of a regulatory risk-based capital program, the insurer may wish to set its own standards.

Other applications for the risk measurement concepts presented here include setting risk loadings for reinsurer default, since the relationship between a ceding insurer and an assuming reinsurer is analogous to that of a policyholder and an insurer (the ceding commission for reinsured business should include a provision for the reinsurer's possible insolvency). Another practical use might be establishing solvency ratings for insurers based on the relationship between their recorded capital (adjusted for known bias) and their risk-based capital.

Although some empirical results have been presented in order to explain the application of the methodology, the findings are still rudimentary. It is especially important to determine more accurate assumptions for the distribution of loss reserve risk and to measure the correlation between risk elements.

Appendix

Expected Policyholder Deficit Under the Normal Distribution

The amount of capital is $\tilde{C} = \tilde{A} - \tilde{L}$, where $A = (1+c)L$. $\tilde{A}$ is the value of assets and $\tilde{L}$ is the value of unpaid losses, both random variables. The expected value of $\tilde{C}$ is $cL \equiv \mu$ and the variance of $\tilde{C}$ is $\sigma^2 = \sigma^2_A - 2\rho\sigma_A\sigma_L + \sigma^2_L$, where σ^2_A and σ^2_L are the respective variances of assets and losses, and ρ is the correlation coefficient. The policyholder deficit is $\tilde{L} - \tilde{A} = -\tilde{C}$ for $\tilde{L} > \tilde{A}$ or $\tilde{C} < 0$. The expected policyholder deficit is $D = \int_{-\infty}^{0} -zp(z)dz$, where $p(z) = [1/(\sigma(\sqrt{2\pi}\,)] \exp\{-(z-\mu)^2/2\sigma^2\}$ is the normal probability density function. Let $y = (z - \mu)/\sigma$. Then $dz = \sigma dy$, and we have

$D = \int_{-\infty}^{-\mu/\sigma} \left[-(\mu+y\sigma)/\sigma\sqrt{2\pi}\right] \exp(-y^2/2)dy$, which reduces to

$$D = \sigma\phi\left(\frac{-\mu}{\sigma}\right) - \mu\Phi\left(\frac{-\mu}{\sigma}\right), \tag{A1}$$

where $\Phi(\bullet)$ = the cumulative standard normal distribution, and $\phi(\bullet)$ is the standard normal density. Notice that the probability of ruin ($\tilde{C} < 0$) is $\Phi(-\mu/\sigma)$. Define $k_T \equiv \sigma/L$, the ratio of the standard deviation of the capital (total assets minus losses) to the expected loss. Then $\mu/\sigma = (cL)/(k_TL) = c/k_T$. The EPD ratio is

$$d = \frac{D}{L} = k_T\phi\left(\frac{-c}{k_T}\right) - c\Phi\left(\frac{-c}{k_T}\right). \tag{A2}$$

Letting the variance of assets be zero, we have the expected policyholder deficit ratio for risky losses:

$$d_L = k\phi\left(\frac{-c}{k}\right) - c\Phi\left(\frac{-c}{k}\right), \tag{A3}$$

where $k = \sigma_L/L$, with σ_L being the standard deviation of losses.

Let $c_A = C/A = c/(1+c)$ be the capital/assets ratio; thus, $c = c_A/(1-c_A)$. The asset coefficient of variation is $k_A = \sigma_A/L(1+c)$. Setting the variance of losses to zero, we get $\sigma = \sigma_A$. Then $k_T = \sigma_A/L = k_A(1+c) = k_A/(1-c_A)$, and we have the EPD ratio for risky assets:

$$d_A = \frac{D_A}{L} = \frac{1}{1-c_A}\left[k_A\phi\left(\frac{-c_A}{k_A}\right) - c_A\Phi\left(\frac{-c_A}{k_A}\right)\right]. \tag{A4}$$

Expected Policyholder Deficit Under the Lognormal Distribution

To determine the lognormal expected policyholder deficit at the end of one period with no time value ($i = 0$), we use the fact that the EPD for risky losses is a call option with exercise price A and current "stock price" L. Since the well-known Black-Scholes (1973) option pricing model assumes that the future stock price is subject to geometric Brownian motion with instantaneous variance σ^2, at time t the price is lognormally distributed with dispersion parameter $\sigma\sqrt{t}$. The option price is

$$F = S\Phi(a) - Ee^{-it}\Phi(a - \sigma\sqrt{t}), \tag{A5}$$

where $a = \dfrac{\ln(S/E) + (i + \sigma^2/2)t}{\sigma\sqrt{t}}$,

S = the stock price, and

E = the exercise price.

Substituting $i = 0$, $t = 1$, $\sigma = \sigma_L$, $A = (1+c)L = E$ and $L = S$, we get the expected policyholder deficit

$$D_L = L\Phi(a) - (1+c)L\Phi(a - \sigma_L), \tag{A6}$$

where $a = (\sigma_L/2) - (\ln(1+c)/\sigma_L)$, and $\Phi(\bullet)$ = the cumulative standard normal distribution. The coefficient of variation is $k = \sqrt{\exp(\sigma_L^2) - 1} \cong \sigma_L$. Let $k = \sigma_L$ (however, for large values of k, we can use the exact relationship). Dividing equation (A6) by L, the EPD ratio to expected loss is

$$d_L = \Phi(a) - (1+c)\Phi(a - k). \tag{A7}$$

For risky assets with dispersion parameter σ_A, the expected policyholder deficit is equivalent to a put option with exercise price L and stock price A. The corresponding call option value is

$$D'_L = A\Phi(a') - L\Phi(a' - \sigma_A), \tag{A8}$$

where (following the preceding derivation, with $\sigma_A = \sigma_L$) we have $a' = (\sigma_A/2) + (\ln(1+c)/\sigma_A) = \sigma_A - a$. Thus, $D'_L = A\Phi(\sigma_A - a) - L\Phi(-a) = A[1 - \Phi(a - \sigma_A)] - L[1 - \Phi(a)] = D_L + A - L$.

To determine the value of the put option, we use the *put-call parity* relationship $G = F' - S + Ee^{-it}$, where G and F′ are the respective values of put and call options with stock price $S = A$ and exercise price $E = L$. Since $i = 0$, the EPD is $G = D_A = D'_L - A + L = D_L$, and the EPD ratio equals d_L. Since $c = c_A/(1-c_A)$ from the normal EPD derivation and we use the approximation $k_A = \sigma_A$, the EPD ratio in equation (A7) can be cast in terms of the risky asset parameters k_A and c_A:

$$d_A = \Phi(b) - \frac{\Phi(b-k_A)}{1-c_A}, \tag{A9}$$

where $b = (k_A/2) + (\ln(1-c_A)/k_A)$.

When both $\tilde{A}$ and $\tilde{L}$ are random variables, the EPD can be determined from the associated put option, where assets behave like the stock price and losses like a stochastic exercise price. Cummins (1988) evaluates this option cost under a continuous time framework using the fact that the variable $\tilde{A}/\tilde{L}$ is also lognormal. It has dispersion parameter σ based on the underlying bivariate normal distribution (see the preceding section) where $\sigma^2 = \sigma^2_A - 2\rho\sigma_A\sigma_L + \sigma^2_L$. The value of this option is the same as that for another insurer with risky assets also having expected value A but with dispersion parameter σ instead of σ_A and riskless losses equal to L. Thus, as with the normal distribution, the total risk (measured by σ) determines the EPD, which in this case can be evaluated by equation (A9).

References

Aitchison, J. and J. A. Brown, 1969, *The Lognormal Distribution* (Cambridge: Cambridge University Press).

A. M. Best Company, 1991, *Best's Insolvency Study, Property Casualty Insurers 1969-1990* (Oldwick, N.J.: A. M. Best).

Beard, R. E., T. Pentikäinen, and E. Pesonen, 1984, *Risk Theory*, Third Edition (London: Chapman and Hall).

Black, Fischer and Myron Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81: 637-659.

Brealy, Richard A. and Stuart C. Myers, 1988, *Principles of Corporate Finance*, Third Edition (New York: McGraw-Hill).

Brockett, Patrick L. and Robert C. Witt, 1990, Relevant Distributions for Insurance Prices in an Arbitrage Free Equilibrium, *Journal of Risk and Insurance*, 58:13-29.

Butsic, Robert P., 1988, Determining the Proper Interest Rate for Loss Reserve Discounting: An Economic Approach, *Evaluating Insurance Company Liabilities*, Casualty Actuarial Society Discussion Paper Program (New York: Casualty Actuarial Society), 147-188.

Byrnes, J. F., 1986, A Survey of the Relationship Between Claims Reserves and Solvency Margins, *Insurance: Mathematics and Finance*, 5: 3-29.

Conning & Company, 1989, *The Real Property & Casualty Underwriting Cycle: Behind the Reported Results* (Hartford: Conn.: Conning & Company).

Cox, John C., Stephen A. Ross, and Mark Rubinstein, 1979, Option Pricing: A Simplified Approach, *Journal of Financial Economics*, 7: 229-263.

Cox, John C. and Mark Rubinstein, 1985, *Options Markets* (Englewood Cliffs, N.J.: Prentice-Hall).

Cummins, J. David, 1988, Risk-Based Premiums for Insurance Guaranty Funds, *Journal of Finance*, 43: 823-838.

D'Arcy, Stephen P., 1988, Use of the CAPM to Discount Property-Liability Loss Reserves, *Journal of Risk and Insurance*, 55: 481-491.

Daykin, C. D., G. D. Bernstein, S. M. Coutts, E. R. F. Devitt, G. B. Hey, D. I. W. Reynolds, and P. D. Smith, 1987, Assessing the Solvency and Financial Strength of a General Insurance Company, *Journal of the Institute of Actuaries*, 114: 227-325.

Derrig, Richard A., 1989, Solvency Levels and Risk Loadings Appropriate for Fully Guaranteed Property-Liability Insurance Contracts: A Financial View, in: J. D. Cummins and R. A. Derrig eds., *Financial Models of Insurance Solvency* (Boston: Kluwer Academic Publishers), 303-354.

Doherty, Neil A. and James R. Garven, 1986, Price Regulation in Property-Liability Insurance: A Contingent-Claims Approach, *Journal of Finance*, 41: 1031-1050.

Garven, James R., 1992, An Exposition of the Implications of Limited Liability and Asymmetric Taxes for Property-Liability Insurance, *Journal of Risk and Insurance*, 59: 34-56.

Ibbotson Associates, 1990, *Stocks, Bonds, Bills and Inflation 1989 Yearbook* (Chicago: Ibbotson Associates).

Samuelson, Paul, 1983, *Foundations of Economic Analysis*, Enlarged Edition (Cambridge, Mass.: Harvard University Press).

Trigeorgis, Lenos, 1991, A Log-Transformed Binomial Numerical Analysis Method for Valuing Complex Multi-Option Investments, *Journal of Financial and Quantitative Analysis*, 26: 309-326.

CAS Exam 9 Study Kit

"The Economics of Structured Finance"

by J. Coval, J. Jurek, and E. Stafford

Journal of Economic Perspectives

Journal of Economic Perspectives—Volume 23, Number 1—Winter 2009—Pages 3–25

The Economics of Structured Finance

Joshua Coval, Jakub Jurek, and Erik Stafford

The essence of structured finance activities is the pooling of economic assets like loans, bonds, and mortgages, and the subsequent issuance of a prioritized capital structure of claims, known as tranches, against these collateral pools. As a result of the prioritization scheme used in structuring claims, many of the manufactured tranches are far safer than the average asset in the underlying pool. This ability of structured finance to repackage risks and to create "safe" assets from otherwise risky collateral led to a dramatic expansion in the issuance of structured securities, most of which were viewed by investors to be virtually risk-free and certified as such by the rating agencies. At the core of the recent financial market crisis has been the discovery that these securities are actually far riskier than originally advertised.

We examine how the process of securitization allowed trillions of dollars of risky assets to be transformed into securities that were widely considered to be safe, and argue that two key features of the structured finance machinery fueled its spectacular growth. First, we show that most securities could only have received high credit ratings if the rating agencies were extraordinarily confident about their ability to estimate the underlying securities' default risks, and how likely defaults were to be correlated. Using the prototypical structured finance security—the collateralized debt obligation (CDO)—as an example, we illustrate that issuing a capital structure amplifies errors in evaluating the risk of the underlying securities.

■ *Joshua Coval is Professor of Business Administration at Harvard Business School, Boston, Massachusetts, and Jakub Jurek is Assistant Professor at Princeton University, Princeton, New Jersey, and Erik Stafford is Associate Professor of Business Administration at Harvard Business School, Boston, Massachusetts. Their e-mail addresses are ⟨jcoval@hbs.edu⟩, ⟨jjurek@princeton.edu⟩, and ⟨estafford@hbs.edu⟩.*

In particular, we show how modest imprecision in the parameter estimates can lead to variation in the default risk of the structured finance securities that is sufficient, for example, to cause a security rated AAA to default with reasonable likelihood. A second, equally neglected feature of the securitization process is that it substitutes risks that are largely diversifiable for risks that are highly systematic. As a result, securities produced by structured finance activities have far less chance of surviving a severe economic downturn than traditional corporate securities of equal rating. Moreover, because the default risk of senior tranches is concentrated in systematically adverse economic states, investors should demand far larger risk premia for holding structured claims than for holding comparably rated corporate bonds. We argue that both of these features of structured finance products—the extreme fragility of their ratings to modest imprecision in evaluating underlying risks and their exposure to systematic risks—go a long way in explaining the spectacular rise and fall of structured finance.

For over a century, agencies such as Moody's, Standard and Poor's, and Fitch have gathered and analyzed a wide range of financial, industry, and economic information to arrive at independent assessments on the creditworthiness of various entities, giving rise to the widely popular rating scales (AAA, AA, A, BBB, and so on). Until recently, the agencies focused the majority of their business on single-name corporate finance—that is, issues of creditworthiness of financial instruments that can be clearly ascribed to a single company. In recent years, the business model of credit rating agencies has expanded beyond their historical role to include the nascent field of structured finance.

From its beginnings, the market for structured securities evolved as a "rated" market, in which the risk of tranches was assessed by credit rating agencies. Issuers of structured finance products were eager to have their new products rated on the same scale as bonds so that investors subject to ratings-based constraints would be able to purchase the securities. By having these new securities rated, the issuers created an illusion of comparability with existing "single-name" securities. This provided access to a large pool of potential buyers for what otherwise would have been perceived as very complex derivative securities.

During the past decade, risks of all kinds have been repackaged to create vast quantities of triple-A–rated securities with competitive yields. By mid-2007, there were 37,000 structured finance issues in the U.S. alone with the top rating (Scholtes and Beales, 2007). According to Fitch Ratings (2007), roughly 60 percent of all global structured products were AAA-rated, in contrast to less than 1 percent of the corporate issues. By offering AAA-ratings along with attractive yields during a period of relatively low interest rates, these products were eagerly bought up by investors around the world. In turn, structured finance activities grew to represent a large fraction of Wall Street and rating agency revenues in a relatively short period of time. By 2006, structured finance issuance led Wall Street to record revenue and compensation levels. The same year, Moody's Corporation reported that 44 percent of its revenues came from rating structured finance products, surpassing the 32 percent of revenues from their traditional business of rating corporate bonds.

By 2008, everything had changed. Global issuance of collateralized debt obligations slowed to a crawl. Wall Street banks were forced to incur massive write-downs. Rating agency revenues from rating structured finance products disappeared virtually overnight and the stock prices of these companies fell by 50 percent, suggesting the market viewed the revenue declines as permanent. A huge fraction of existing products saw their ratings downgraded, with the downgrades being particularly widespread among what are called "asset-backed security" collateralized debt obligations—which are comprised of pools of mortgage, credit card, and auto loan securities. For example, 27 of the 30 tranches of asset-backed collateralized debt obligations underwritten by Merrill Lynch in 2007 saw their triple-A ratings downgraded to "junk" (Craig, Smith, and Ng, 2008). Overall, in 2007, Moody's downgraded 31 percent of all tranches for asset-backed collateralized debt obligations it had rated and 14 percent of those initially rated AAA (Bank of International Settlements, 2008). By mid-2008, structured finance activity was effectively shut down, and the president of Standard & Poor's, Deven Sharma, expected it to remain so for "years" (*Financial Week,* 2008).

This paper investigates the spectacular rise and fall of structured finance. We begin by examining how the structured finance machinery works. We construct some simple examples of collateralized debt obligations that show how pooling and tranching a collection of assets permits credit enhancement of the senior claims. We then explore the challenge faced by rating agencies, examining, in particular, the parameter and modeling assumptions that are required to arrive at accurate ratings of structured finance products. We then conclude with an assessment of what went wrong and the relative importance of rating agency errors, investor credulity, and perverse incentives and suspect behavior on the part of issuers, rating agencies, and borrowers.

Manufacturing AAA-rated Securities

Manufacturing securities of a given credit rating requires tailoring the cash-flow risk of these securities—as measured by the likelihood of default and the magnitude of loss incurred in the event of a default—to satisfy the guidelines set forth by the credit rating agencies. Structured finance allows originators to accomplish this goal by means of a two-step procedure involving pooling and tranching.

In the first step, a large collection of credit-sensitive assets is assembled in a portfolio, which is typically referred to as a "special purpose vehicle." The special purpose vehicle is separate from the originator's balance sheet to isolate the credit risk of its liabilities—the tranches—from the balance sheet of the originator. If the special purpose vehicle issued claims that were not prioritized and were simply fractional claims to the payoff on the underlying portfolio, the structure would be known as a pass-through securitization. At this stage, since the expected portfolio loss is equal to the mean expected loss on the underlying securities, the portfolio's credit rating would be given by the average rating of the securities in the underlying pool. The pass-through securitization claims would inherit this rating, thus achieving no credit enhancement.

By contrast, to manufacture a range of securities with different cash flow risks, structured finance issues a capital structure of prioritized claims, known as *tranches*, against the underlying collateral pool. The tranches are prioritized in how they absorb losses from the underlying portfolio. For example, senior tranches only absorb losses after the junior claims have been exhausted, which allows senior tranches to obtain credit ratings in excess of the average rating on the average for the collateral pool as a whole. The degree of protection offered by the junior claims, or overcollateralization, plays a crucial role in determining the credit rating for a more senior tranche, because it determines the largest portfolio loss that can be sustained before the senior claim is impaired.

This process of pooling and tranching, common to all structured securities, can be illustrated with a two-asset example. Consider two identical securities—call them "bonds"—both of which have a probability of default p_D, and pay \$0 conditional on default and \$1 otherwise. Suppose we pool these securities in a portfolio, such that the total notional value of the underlying fund is \$2, and then issue two tranches against this fund, each of which pay \$1. A "junior" tranche can be written such that it bears the first \$1 of losses to the portfolio; thus, the junior tranche pays \$1 if both bonds avoid default and zero if either bond defaults. The second, "senior" claim, which bears losses if the capital of the junior tranche is exhausted, pays \$1 if neither bond defaults or if only one out of two bonds defaults; it only defaults if both bonds default. It should be intuitively clear that to compute the expected cash flows (or default probabilities) for the tranches, we will need to know the likelihood of observing both bonds defaulting simultaneously. In this example, the default dependence structure can be succinctly described by means of a single parameter—either the joint probability of default, or the default correlation.[1]

What makes this structure interesting is that if the defaults of the two bonds are imperfectly correlated, the senior tranche will pay either \$1 or \$0—just like the individual bonds—except that it will be less likely to default than either of the underlying bonds. For example, if the two bonds have a 10 percent default probability and defaults are uncorrelated, the senior tranche will only have a 1 percent chance of default. This basic procedure allows highly risky securities to be repackaged, with some of the resulting tranches sold to investors seeking only safe investments. Obviously, junior tranches, being risky, will have low prices and high

[1] If we assume that both securities are identical and denote the probability of observing both claims default simultaneously by p_{DD}, the *default correlation* parameter can be computed as $(p_{DD} - p_D^2)/(p_D{*}(1 - p_D)$.

promised returns, while the senior tranches, being relatively safe, will have relatively higher prices and lower promised returns.

A central insight of structured finance is that by using a larger number of securities in the underlying pool, a progressively larger fraction of the issued tranches can end up with higher credit ratings than the average rating of the underlying pool of assets. For example, consider extending the two-bond example by adding a third $1 bond, so that now three $1 claims can be issued against this underlying capital structure. Now, the first tranche defaults if any of the three bonds default, the second tranche defaults if two or more of the bonds default, and the final, senior-most tranche only defaults when all three bonds default. If bonds default 10 percent of the time and defaults are uncorrelated, the senior tranche will now default only 0.1 percent of the time, the middle tranche defaults 2.8 percent of the time, and the junior tranche defaults 27.1 percent of the time. Thus, by including a third bond in the pool, two-thirds of the capital—as measured by the tranche notional values—can be repackaged into claims that are less risky than the underlying bonds.

Another way to increase the total notional value of highly-rated securities produced is to reapply the securitization machinery to the junior tranches created in the first round. For example, in the two-bond case in which defaults were uncorrelated, the $1 junior tranche defaults with 19 percent probability. However, if we combine this $1 junior tranche with an identical $1 junior tranche created from another two-bond pool, we can again tranche the resulting $2 of capital into two prioritized $1 claims. If there continues to be no correlation among underlying assets, the resulting senior tranche from this second round of securitization—a tranche that defaults if at least one bond defaults in *each* of the two underlying pools—has a default probability of 3.6 percent, which is once again considerably lower than that of the underlying bonds. The collateralized debt obligations created from the tranches of other collateralized debt obligations are typically called CDO-squared—that is, CDO^2.

A key factor determining the ability to create tranches that are safer than the underlying collateral is the extent to which defaults are correlated across the underlying assets. The lower the default correlation, the more improbable it is that all assets default simultaneously and therefore the safer the senior-most claim can be made. Conversely, as bond defaults become more correlated, the senior-most claims become less safe. Consider, for example, the two-bond case in which defaults are perfectly correlated. Since now both bonds either survive or default simultaneously, the structure achieves no credit enhancement for the senior tranche. Thus, in the two-bond example, while uncorrelated risks of default allow the senior claim to have a 1 percent default probability, perfectly correlated risks of default would mean that the senior claim inherits the risk of the underlying assets, at 10 percent. Finally, intermediate levels of correlation allow the structure to produce a senior claim with default risk between 1 and 10 percent.

The Challenge of Rating Structured Finance Assets

Credit ratings are designed to measure the ability of issuers or entities to meet their future financial commitments, such as principal or interest payments. Depending on the agency issuing the rating and the type of entity whose creditworthiness is being assessed, the rating is either based on the anticipated likelihood of observing a default, or it is based on the expected economic loss—the product of the likelihood of observing a default and the severity of the loss conditional on default. As such, a credit rating can intuitively be thought of as a measure of a security's expected cash flow.[2] In the context of corporate bonds, securities rated BBB− or higher have come to be known as *investment grade* and are thought to represent low to moderate levels of default risk, while those rated BB+ and below are referred to as *speculative grade* and are already in default or closer to it.

Table 1 reports Fitch's estimates regarding the 10-year default probabilities of corporate bonds with different ratings at issuance and gives their corresponding annualized default rates. These estimates are derived from a study of historical data and are used in Fitch's model for rating collateralized debt obligations (Derivative Fitch, 2006).[3] It is noteworthy that within the investment grade range, there are ten distinct rating categories (from AAA to BBB−) even though the annualized default rate only varies between 0.02 and 0.75 percent. Given the narrow range of the historical default rates, distinguishing between the ratings assigned to investment grade securities requires a striking degree of precision in estimating a security's default likelihood. By contrast, the ten rating categories within the speculative grade range (from BB+ to C) have default rates ranging from 1.07 to 29.96 percent.

In the single-name rating business, where the credit rating agencies had developed their expertise, securities were assessed independently of each other, allowing rating agencies to remain agnostic about the extent to which defaults might be correlated. But to assign ratings to structured finance securities, the rating agencies were forced to address the bigger challenge of characterizing the entire *joint* distribution of payoffs for the underlying collateral pool. As the previous section demonstrated, the riskiness of collateralized debt obligation tranches is sensitive to the extent of commonality in default among the underlying assets, since collateralized debt obligations rely on the power of diversification to achieve credit enhancement.

The structure of collateralized debt obligations magnifies the effect of impre-

[2] Credit rating agencies stress that their ratings are only designed to provide an ordinal ranking of securities' long-run ("through-the-cycle") payoff prospects, whereas the expected cash flow interpretation takes a cardinal view of ratings.

[3] A comprehensive description of Fitch's rating model for collateralized debt obligations—the Default VECTOR Model—including assumptions regarding default probabilities, recovery rates, and correlations is available online. An Excel spreadsheet implementation of the model can be downloaded from ⟨http://www.fitchrating.com/jsp/corporate/ToolsAndModels.faces?context=2&detail=117⟩.

Table 1
Historical Default Experience of Bonds Rated by Fitch

	Investment-grade Bonds									
Rating at issuance	*AAA*	*AA+*	*AA*	*AA−*	*A+*	*A*	*A−*	*BBB+*	*BBB*	*BBB−*
10-year default probability	0.19%	0.57%	0.89%	1.15%	1.65%	1.85%	2.44%	3.13%	3.74%	7.26%
Default rate (annualized)	0.02%	0.06%	0.09%	0.12%	0.17%	0.19%	0.25%	0.32%	0.38%	0.75%

	Speculative-grade Bonds									
Rating at issuance	*BB+*	*BB*	*BB−*	*B+*	*B*	*B−*	*CCC+*	*CCC*	*CC*	*C*
10-year default probability	10.18%	13.53%	18.46%	22.84%	27.67%	34.98%	43.36%	48.52%	77.00%	95.00%
Default rate (annualized)	1.07%	1.45%	2.04%	2.59%	3.24%	4.30%	5.68%	6.64%	14.70%	29.96%

cise estimates of default likelihoods, amounts recovered in the event of default, default correlation, as well as model errors due to the potential misspecification of default dependencies (Tarashev and Zhu, 2007; Heitfield, 2008). These problems are accentuated further through the sequential application of capital structures to manufacture the collateralized debt obligation (CDO) tranches commonly known as CDO^2. With multiple rounds of structuring, even minute errors at the level of the underlying securities that would be insufficient to alter the security's rating can dramatically alter the ratings of the structured finance securities.

To illustrate the sensitivity of the collateralized debt obligations and their progeny, the CDO^2, to errors in parameter estimates, we conduct a simulation exercise. First, we simulate the payoffs to 40 collateralized debt obligation pools, each comprised of 100 bonds with a five-year default probability of 5 percent and a recovery rate of 50 percent of face value conditional on default.[4] Using the annualized default rates reported in Table 1 as a guide, each bond in our hypothetical collateral pool would garner a just-below investment grade rating of BB+. Finally, we fix the pairwise bond default correlation at 0.20 within each collateral pool and assume the defaults of bonds belonging to different collateral pools are uncorrelated. Our simulation methodology relies upon a simplified version of the model that is the industry standard for characterizing portfolio losses.[5]

[4] Recovery rates can vary by type of security, seniority, and the country of origin. Historical recovery rates are between 40 and 50 percent for senior unsecured corporate bonds in the United States (Derivative Fitch, 2006; Altman, 2006).

[5] The common method for modeling the joint incidence of defaults is known as the copula method (Schonbucher, 2003). This approach draws a set of N correlated random variables $\{X_i\}$ from a pre-

Within each collateral pool, we construct a capital structure comprised of three tranches prioritized in order of their seniority. The "junior tranche" is the first to absorb losses from the underlying collateral pool and does so until the portfolio loss exceeds 6 percent, at which point the junior tranche becomes worthless. The "mezzanine tranche" begins to absorb losses once the portfolio loss exceeds 6 percent and continues to do so until the portfolio loss reaches 12 percent. Finally, the senior tranche absorbs portfolio losses in excess of 12 percent. We also construct a CDO^2—to be called "CDO^2 [6–12]"—by issuing a second capital structure of claims against a pool that combines the mezzanine tranches from the 40 original collateralized debt obligations.

While the parameter values used in our simulation do not map into any particular market, they were chosen to mimic broadly the types of collateral and securitizations commonly observed in structured finance markets.[6] After simulating the payoffs to the underlying collateral, our first step is to assign ratings to the tranches. We do this by comparing the simulated likelihood of impairment to each tranche's capital with the five-year default probability based on the annualized default rates reported in Table 1. Under our baseline parameters, the mezzanine tranche of the original collateralized debt obligation garners the lowest investment grade rating of BBB−, while the senior tranche—accounting for 88 percent of capital structure—receives a AAA rating. The collateralized debt obligation made up of mezzanine tranches, "CDO^2 ([6, 12])" in the bottom panel of Table 2, has mezzanine and senior tranches that are able to achieve a rating of AAA. Table 2 describes the default probabilities and expected payoffs (as a fraction of notional value) for the simulated tranches of both the original collateralized debt obligation and of the CDO^2 constructed from the mezzanine tranches.

Of course, these estimates of risk depend crucially on whether default correlations have been estimated correctly. Figure 1 explores the sensitivity of the

---

specified distribution and then assumes that a firm defaults if its variable, $X_i = x_i$, is below the p-th percentile of the corresponding marginal distribution, $F_i(x_i)$. Under this scheme, by construction, a firm defaults p percent of the time and default dependence can be flexibly captured through the proposed joint distribution for $\{X_i\}$. A popular choice for the joint distribution function is the multivariate Gaussian (Vasicek, 2002), in which default correlation is simply controlled by the pairwise correlation of (X_i, X_j). Popular off-the-shelf CDO rating toolkits offered by credit rating agencies, such as Fitch's Default VECTOR models, Moody's CDOROM, and Standard and Poor's CDO Evaluator, all employ versions of this copula model.

[6] For example, collateralized loan obligations tend to be issued in a three-tranche structure with attachment points of 0–5 percent, 5–15 percent, and 15–100 percent. Collateralized debt obligations referencing a commonly used index of credit default swaps on corporate bonds have a more granular capital structure with two types of junior claims (0–3 percent and 3–7 percent), two types of mezzanine claims (7–10 percent and 10–15 percent), and two types of senior claims (15–30 percent and 30–100 percent). Tranches that are based on an index of residential mortgage-backed securities have a similarly granular structure with junior claims having attachment points of 0–3 percent and 3–7 percent; mezzanine claims, 7–12 percent and 12–20 percent; and senior claims, 20–35 percent and 35–100 percent.

Table 2
Summary Statistics for CDO and CDO² Tranches in our Simulation under Baseline Parameters

	Attachment points	*Default probability*	*Expected payoff*	*Rating*
CDO				
Junior	0%–6%	97.52%	0.59	NR
Mezzanine	6%–12%	2.07%	> 0.99	BBB–
Senior	12%–100%	< 0.00%	> 0.99	AAA
CDO^2 ([6, 12])				
Junior	0%–6%	56.94%	0.93	C
Mezzanine	6%–12%	< 0.00%	> 0.99	AAA
Senior	12%–100%	< 0.00%	> 0.99	AAA

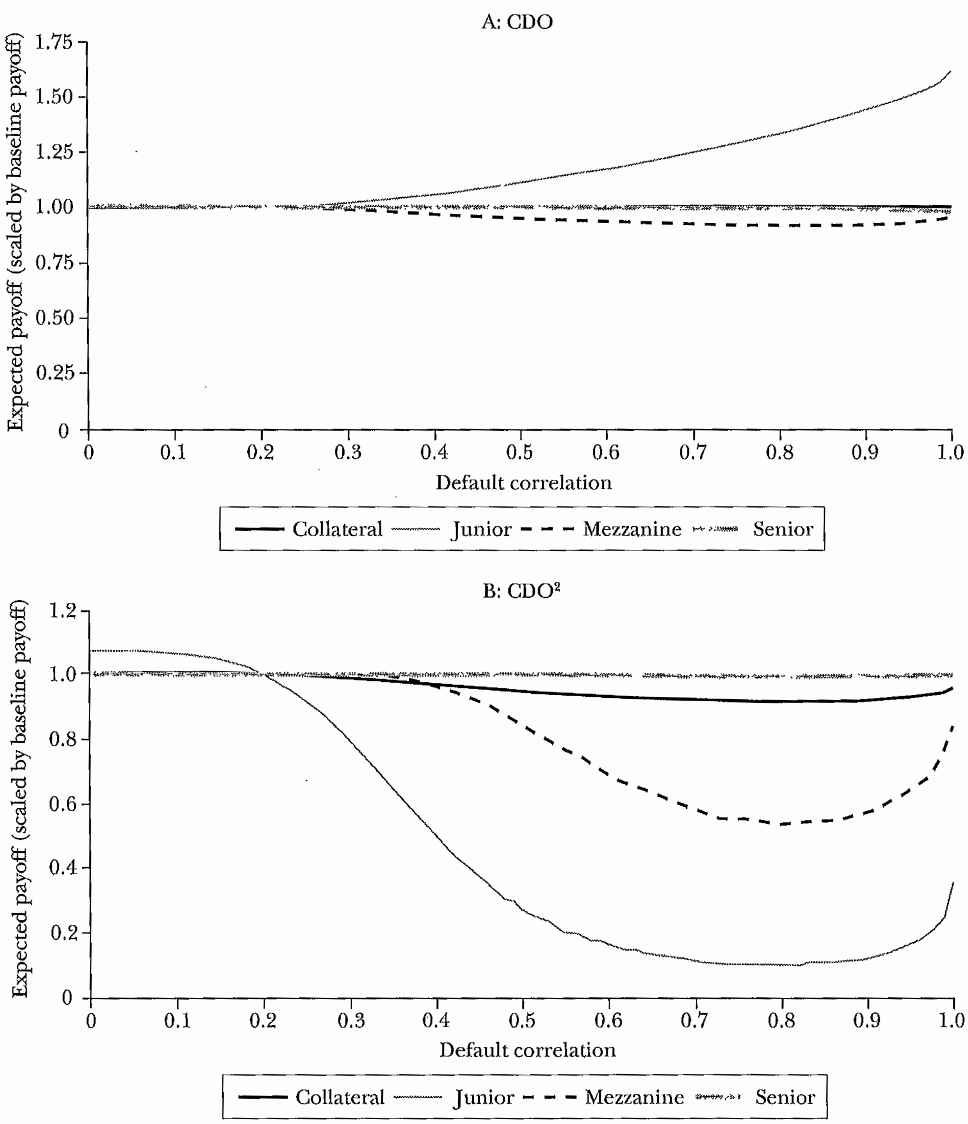
Note: While the parameter values used in our simulation do not map into any particular market, they were chosen to mimic broadly the types of collateral and securitizations commonly observed in structured finance markets.

original collateralized debt obligation and the CDO^2 tranches to changes in default correlation for bonds *within* each collateralized debt obligation. The correlation in defaults for bonds belonging to different collateral pools remains fixed at zero. The figure displays the expected payoff as a function of the default correlation, normalized by the expected payoff under the baseline calibration. These values can be thought of as illustrating the impact of either an error in the modeling assumptions or an unexpected realization of the default experience on the value of a $1 investment in each tranche.

The top panel shows that the expected payoff of the underlying collateral pool does not depend on the default correlation. As the default correlation increases from its baseline value of 0.20, indicating default risk is less diversified than expected, risk shifts from the junior claims to the senior claims. Consequently, the expected payoff on the junior tranche rises relative to the baseline value, while the expected payoff on the mezzanine tranche falls. The effect of changes in default correlation on the mezzanine tranche of the collateralized debt obligation is nonmonotonic. The expected payoff declines until the default correlation reaches a value of 0.80, where the tranche has lost approximately 10 percent of its value relative to the baseline calibration, and then rises as defaults become perfectly correlated and risk is shifted toward the senior tranche. In the limit of perfect default correlation, each tranche faces the same 5 percent chance of default over five years as we assigned each of the individual securities in the underlying portfolio.

The bottom panel of Figure 1 shows how shifts in the valuation of the mezzanine tranche of the collateralized debt obligation are amplified by the second-generation capital structure of the CDO^2. For example, as the pairwise default correlations within the underlying collateral pool of bonds increase from

Figure 1
Sensitivity of CDO and CDO^2 to Changes in Default Correlation



Note: Figure 1 explores the sensitivity of the original collateralized debt obligation and the CDO^2 tranches to changes in default correlation for bonds *within* each collateralized debt obligation. The correlation in defaults for bonds belonging to different collateral pools remains fixed at zero. The figure displays the expected payoff as a function of the default correlation, normalized by the expected payoff under the baseline calibration.

20 to 60 percent, the expected payoff on the mezzanine claim of the CDO^2, which is an investment grade security under the baseline parameters, drops by a staggering 25 percent.

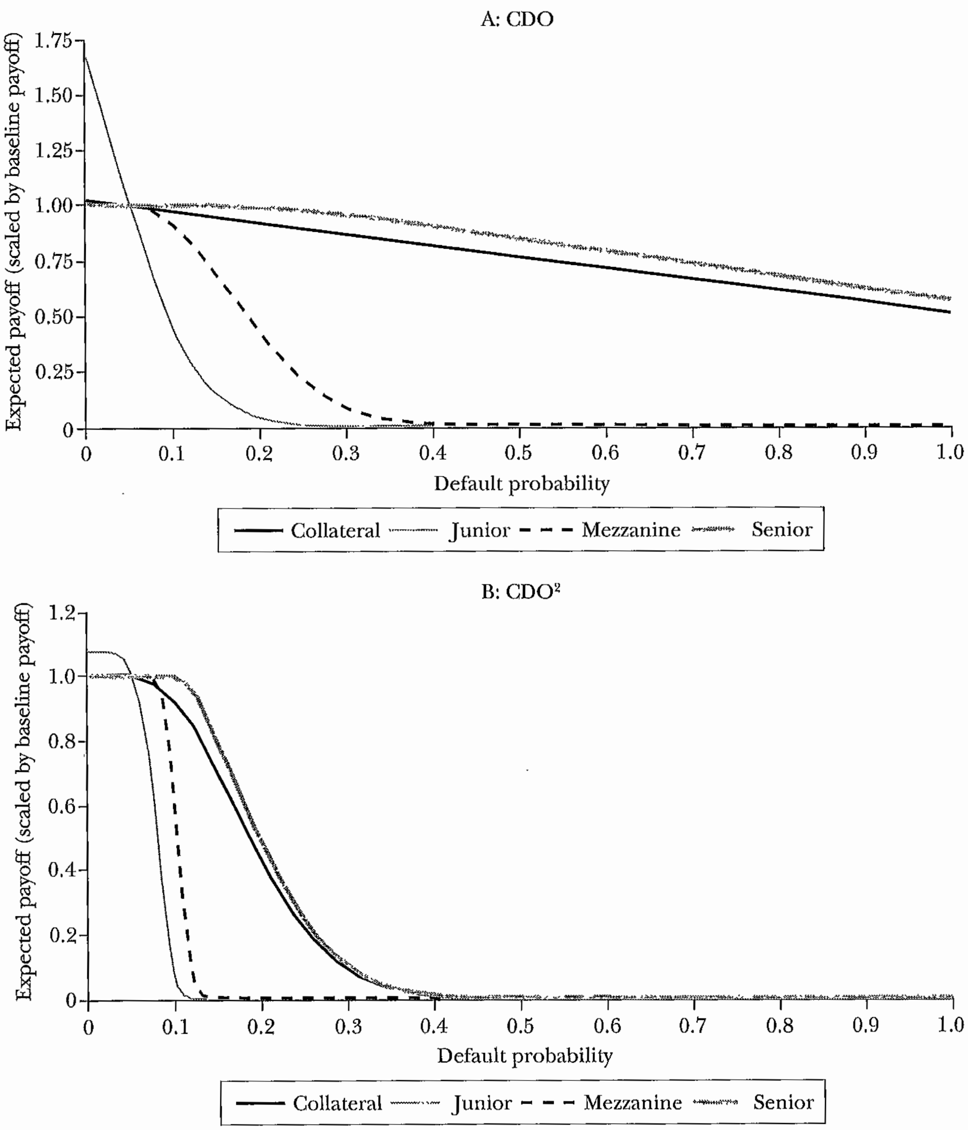
In Figure 2, we examine the effect of errors in estimates of the probability of default on the underlying securities on the expected tranche payoffs, while holding default correlation fixed at the baseline value of 0.20. As the default probability increases (declines) relative to the baseline estimate of 5 percent, the expected payoff on the underlying collateral decreases (increases) monotonically, and this effect is transferred to the tranches of the collateralized debt obligation. The sensitivity of the tranches to errors in the estimate of default probability is determined by their seniority. For example, an increase in the default probability from 5 to 10 percent results in a 55 percent decline in the expected payoff for the junior tranche, an 8 percent decline for the mezzanine tranche, and a 0.01 percent decline for the senior tranche.

The bottom panel of Figure 2 again illustrates the theme that changing the baseline parameters has a much starker effect on the CDO^2 comprised of the mezzanine tranches from the original collateralized debt obligations. In this case, as default probabilities rise, the values of the junior and mezzanine tranches quickly fall towards zero, and the value of the senior tranche falls substantially as well.

Table 3 provides a complementary illustration of how ratings are affected by changes in the underlying assets' default correlation and default probabilities. Although the expected payoff of the senior tranche of the collateralized debt obligation is relatively robust to changes in the model parameters, this is somewhat deceiving. Due to the fine partitioning of investment grade ratings, even modest changes in the model parameters can precipitate a meaningful rating downgrade for the senior tranche. For example, the rating of the senior tranche for the original collateralized debt obligation drops to A+ when the default probability reaches 10 percent and reaches the investment grade boundary of BBB− when the default probability reaches 20 percent. Again, the CDO^2 structure significantly amplifies the variation in the expected payoffs. When the default probability is increased to 10 percent, the mezzanine claim of the CDO^2, which was initially rated AAA, sees 50 percent of its expected payoff wiped out and its rating drop all the way below the rating scale. Even a slight increase in the probability of default on the underlying securities to 7.5 percent, which would only cause the underlying securities to be downgraded from BB+ to BB−, is sufficient to precipitate a downgrade of the AAA-rated mezzanine CDO^2 claim to BBB−. Given the plausible uncertainty in estimates of the underlying model parameters, the ".SF" rating modifiers recently proposed by regulators for structured finance instruments (U.S. Securities and Exchange Commission, 2008; Securities Industry and Financial Markets Association, 2008), are perhaps best regarded as warning labels.

Finally, the simulation illustrates that with plausible magnitudes of overcollateralization (the degree of protection offered by the junior claims—12 percent in our example), the expected payoff on a senior tranche of the original

Figure 2.
Sensitivity of CDO and CDO² to Changes in Default Probability



Note: Figure 2 explores the sensitivity of the original collateralized debt obligation and the CDO^2 tranches to changes in the default probability for bonds in each collateralized debt obligation. The figure displays the expected payoff as a function of the default probability, normalized by the expected payoff under the baseline calibration.

collateralized debt obligation is well protected from large changes in default probabilities and correlations. While its rating might change, substantial impairments to the value of such claims seem implausible, short of an economic

Table 3
Effect of Changes in Underlying Parameters on CDO and CDO^2 Tranche Ratings

	Initial rating	*Final rating*					
		Default correlation (ρ)			*Default probability (pD)*		
	(ρ = 20%, pD = 5%)	*40%*	*60%*	*80%*	*7.50%*	*10%*	*12.50%*
CDO							
Junior	NR	D	C	CC	NR	NR	NR
Mezzanine	BBB−	BB−	B+	B+	B+	CCC	CC
Senior	AAA	A+	BBB−	BB	AAA	A+	BBB−
CDO^2 ([6, 12])							
Junior	C	D	NR	NR	NR	NR	NR
Mezzanine	AAA	B+	C	CC	BBB−	NR	NR
Senior	AAA	AAA	AAA	AA+	AAA	AAA	B−

catastrophe. On the other hand, all tranches of the second generation securitization, the CDO^2, are highly sensitive to changes in the baseline parameters. Even slight changes in default probabilities and correlations can have a substantial impact on the expected payoffs and ratings of the CDO^2 tranches, including the most senior claims.

As we show in the next section, a large fraction of collateralized debt obligations issued over the course of the last decade had subprime residential mortgage-backed securities as their underlying assets. Importantly, many of these residential mortgage-backed securities are themselves tranches from an original securitization of a large pool of mortgages, such that collateralized debt obligations of mortgage-backed securities are effectively CDO^{2}s. Moreover, since substantial lending to subprime borrowers is a recent phenomenon, historical data on defaults and delinquencies of this sector of the mortgage market is scarce. The possibility for errors in the assessment of the default correlations, the default probabilities, and the ensuing recovery rates for these securities was significant. Such errors, when magnified by the process of re-securitization, help explain the devastating losses some of these securities have experienced recently.

The Relation of Structured Finance to Subprime

To ensure a continuous supply of credit to home buyers, government-sponsored agencies such as Fannie Mae, Freddie Mac, and Ginnie Mae were chartered to purchase mortgages originated by local banks, provided they satisfy certain size and credit quality requirements. Mortgages conforming to these requirements are repackaged by these agencies into mortgage-backed securities, and resold in capital markets with the implicit guarantee of the U.S. government. In contrast, mortgages that do not conform to size restrictions or borrower credit quality standards are not eligible for purchase by the government-sponsored enterprises and are either held by their issuers or sold directly in secondary markets.[7] In recent years, issuance of so-called "non-conforming" mortgages has increased significantly. For example, origination of subprime mortgages—mortgages given to those below the credit standards for the government-sponsored enterprises—grew from $96.8 billion in 1996 to approximately $600 billion in 2006, accounting for 22 percent of all mortgages issued that year (U.S. Securities and Exchange Commission, 2008). During the same period, the average credit quality of subprime borrowers decreased along a number of measures, as evidenced by rising ratios of mortgage values relative to house prices, an increased incidence of second lien loans, and issuance of mortgages with low or no documentation (Ashcraft and Schuermann, 2008). When house prices declined, the stage was set for a significant increase in default rates as many of these borrowers found themselves holding mortgages in excess of the market value of their homes.

Because subprime mortgages were ineligible for securitization by government-sponsored agencies, they found their way into capital markets by way of "private-label" mortgage-backed securities, originated by Wall Street banks among others (Federal Deposit Insurance Corporation, 2006). These securities carried the dual risk of high rates of default due to the low credit quality of the borrowers; and high levels of default correlation as a result of pooling mortgages from similar geographic areas and vintages. In turn, many subprime mortgage-backed bonds were themselves re-securitized into what are called collateralized mortgage obligations, effectively creating a CDO^2. According to Moody's, the share of collateralized debt obligations that had other "structured" assets as their collateral expanded from 2.6 percent in 1998 to 55 percent in 2006 as a fraction of the total notional value of all securitizations. In 2006 alone, issuance of structured finance collateralized debt obligations reached $350 billion in notional value (Hu, 2007).

As it turned out, *all* of the factors determining expected losses on tranches of collateralized debt obligations backed by mortgage-backed securities had been biased against the investor. First, the overlap in geographic locations and vintages within mortgage pools raised the prospect of higher-than-expected default correlations. Second, the probability of default and the expected recovery values have both been worse than expected due to the deterioration in credit quality of subprime borrowers and because of assets being sold off under financial pressure in "fire sales," further driving down the prices of related assets. Finally, the preva-

[7] Jumbo mortgages have notional values exceeding the conventional loan limit, which was $417,000 for a single-family home in 2008. Subprime borrowers are defined as those with a FICO credit score below 620, limited credit history, or some other form of credit impairment. Alt-A borrowers have credit scores sufficient to qualify for a conforming mortgage, but do not have the necessary documentation to substantiate that their assets and income can support the requested loan amount.

lence of CDO^2 structures further magnified the deleterious effects of errors in estimates of expected losses on the underlying mortgages for investors.

A succinct view of the severity of the deterioration in private-label residential mortgage-backed securities is provided by the ABX.HE indices. These indices are compiled by Markit in cooperation with major Wall Street banks and track the performance of subprime residential mortgage-backed securities along various points in the rating spectrum.[8] For example, the ABX.HE.BBB 07-01 captures the average value of 20 BBB-rated mortgage-backed securities obtained by pooling and tranching subprime mortgages issued in the first half of 2007. Intuitively, each of the underlying mortgage-backed securities can be thought as loosely corresponding to a mezzanine tranche of a collateralized debt obligation in our simulation. Although the ABX.HE.BBB 07-01 index traded as high as 98.35, by August 2008, it had an average rating of CCC and a market price of roughly 5 cents on the dollar. With such abysmal performance in the residential mortgage-backed market, collateralized debt obligations backed by this type of structured collateral are virtually guaranteed to fail. As illustrated by our simulation, a collateralized debt obligation made up of investment grade mezzanine tranches—that is, a CDO^2—can sustain very large losses even with small changes in the realized default probabilities and correlations.

The Pricing of Systematic Risk in Structured Products

When credit rating agencies started rating both structured finance and single-name securities on the same scale, it may well have lured investors seeking safe investments into the structured finance market, even though they did not fully appreciate the nature of the underlying economic risks. In the logic of the capital asset pricing model, securities that are correlated with the market as a whole should offer higher expected returns to investors, and hence have higher yields, than securities with the same expected payoffs (or credit ratings) whose fortunes are less correlated with the market as a whole. However, credit ratings, by design, only provide an assessment of the risks of the security's expected payoff, with no information regarding whether the security is particularly likely to default at the same time that there is a large decline in the stock market or that the economy is in a recession.

Because credit ratings only reflect expected payoffs, securities with a given credit rating can, in theory, command a wide range of yield spreads—that is, yield in excess of the yield on a U.S. Treasury security of the same duration—depending on their exposure to systematic risks. For example, consider a security whose default likelihood is constant and independent of the economic state, such that its

[8] Additional information on the Markit ABX indices, including pricing, can be found at ⟨http://www.markit.com/information/products/category/indices/abx.html⟩.

payoff is unrelated to whether the economy is in a recession or boom, whether interest rates are rising or falling, or the behavior of any other set of economic indicators. An example of this type of a security is a traditional catastrophe bond. Catastrophe bonds are typically issued by insurers and deliver their promised payoff unless there is a natural disaster, such as a hurricane or earthquake, in which case the bond defaults. Under the working assumption that a single natural disaster cannot have a material impact on the world economy, a traditional catastrophe bond will earn a yield spread consistent with compensation for expected losses. Investors are willing to pay a relatively high price for catastrophe bonds because their risks are uncorrelated with other economic indicators and therefore can be eliminated through diversification.

At the other end of the range, the maximum yield spread for a security of a given rating is attained by a security whose defaults are confined to the worst possible economic states. If we assume that the stock market provides an ordering of economic states—that is, if the Standard and Poor's 500 index is at 800, the economy is in worse condition than if that same index is at 900—then the security with maximal exposure to systematic risk is a digital call option on the stock market. A digital call option pays $1 if the market is above a pre-determined level (called a "strike price") at maturity and $0 otherwise. Because this security "defaults" and fails to pay only when the market is below the strike price, it represents the security with the greatest possible exposure to systematic risk. By selecting the appropriate strike price, the probability that the call fails to make its promised payment can be tuned to match any desired credit rating. However, because a digital call option concentrates default in the worst economic states, investors will insist on receiving a high return as compensation for bearing the systematic risk and require the option to deliver the largest yield spread of all securities with that credit rating.

The process of pooling and tranching effectively creates securities whose payoff profiles resemble those of a digital call option on the market index. Intuitively, pooling allows for broad diversification of idiosyncratic default risks, such that—in the limit of a large diversified underlying portfolio—losses are driven entirely by the systematic risk exposure. As a result, tranches written against highly diversified collateral pools have payoffs essentially identical to a derivative security written against a broad economic index.

In effect, structured finance has enabled investors to write insurance against large declines in the aggregate economy. Investors in senior tranches of collateralized debt obligations bear enormous systematic risk, as they are increasingly likely to experience significant losses as the overall economy or market goes down. Such a risk profile should be expected to earn a higher rate of return than those available from single-name bonds, whose defaults are affected by firm-specific bad luck. If investors in senior claims of collateralized debt obligations do not fully appreciate the nature of the insurance they are writing, they are likely to be earning a yield that appears attractive relative to that of securities with similar credit ratings (that is, securities with a similar likelihood of default), but well below the return they could have earned from simply writing such insurance directly—say, by making the appropriate investment in options on the broader stock market index. In Coval, Jurek, and Stafford (forthcoming), we provide evidence for this conjecture, showing that senior tranches in collateralized debt obligations do not offer their investors nearly large enough of a yield spread to compensate them for the actual systematic risks that they bear.

The fact that corporate bonds and structured finance securities carry risks that can, both in principle and in fact, be so different from a pricing standpoint casts significant doubt on whether corporate bonds and structured finance securities can really be considered comparable, regardless of what the credit rating agencies may choose to do.

The Rise and Fall of the Structured Finance Market

The dramatic rise and fall of structured finance products has been remarkable. In under a decade, issuance of these products within the U.S. economy grew more than ten-fold. In the first three quarters of 2005, \$25–\$40 billion of structured finance products were issued in each quarter, according to data from the Securities Industry and Financial Markets Association. In the last quarter of 2006 and the first two quarters of 2007, issuance of structured finance products peaked at about \$100 billion in each quarter. But by the first two quarters of 2008, these quantities had dropped to less than \$5 billion per quarter.

It is easy to see how the events of 2007 and 2008 compelled investors to reassess the risks they were bearing in structured products. Less obvious is how structured finance achieved such amazing growth in such a short period of time. Why were investors eager to purchase structured products and issuers eager to supply them? As we have argued, the key to understanding the market's dramatic rise and fall is to recognize the tendency of pooling and tranching to amplify mistakes in the assessment of underlying asset default risks and correlations as well as their ability to concentrate systematic risks in the most senior tranches.

The rapid growth of the market for structured products coincided with fairly strong economic growth and few defaults, which gave market participants little reason to question the robustness of these products. In fact, all parties believed they were getting a good deal. Many of the structured finance securities with AAA-ratings offered yields that were attractive relative to other, rating-matched alternatives, such as corporate bonds. The "rated" nature of these securities, along with their yield advantage, engendered significant interest from investors. However, these seemingly attractive yields were in fact too low given the true underlying risks. First, the securities' credit ratings provided a downward-biased view of their actual default risks, since they were based on the credit rating agencies' naïve extrapolation of the favorable economic conditions. Second, the yields failed to account for the extreme exposure of structured products to declines in aggregate economic conditions (in other words, systematic risk). The spuriously low yields on senior claims, in turn, allowed the holders of remaining claims to be overcompensated, incentivizing market participants to hold the "toxic" junior tranches. As a result of this mispricing, demand for structured claims of all seniorities grew explosively. The banks were eager to play along, collecting handsome fees for origination and structuring. Ultimately, the growing demand for the underlying collateral assets lead to an unprecedented reduction in the borrowing costs for homeowners and corporations alike, fueling the real estate bubble that is now unwinding.

It seems that few investors were worried that the underlying assets were overvalued, and those who were had incentives to disregard this possibility. This changed rapidly when subprime mortgage defaults started increasing. As we demonstrated earlier, errors in default probabilities adversely affect all of the tranches, with the junior tranches taking the first losses. Moreover, the CDO^2 structure, which was especially common in this market, magnifies these errors, such that even their senior-most tranches can be significantly impaired.

It is tempting to lay the bulk of the blame for the rise and fall of structured finance on the credit rating agencies, since it was the agencies that evaluated and deemed assets created by collateralized debt obligations as "safe." There is certainly evidence that the rating agencies made some significant mistakes. For example, in May 2008, Moody's acknowledged that it had inadvertently given AAA-ratings to billions of dollars of structured finance products due to a bug in one of its ratings models (Jones, Tett, and Davies, 2008). In March 2007, First Pacific Advisors discovered that Fitch used a model that assumed constantly appreciating home prices, ignoring the possibility that they could fall. Robert Rodriguez (2007), the chief executive officer of First Pacific Advisors, describes the discovery:

> We were on the March 22 call with Fitch regarding the sub-prime securitization market's difficulties. In their talk, they were highly confident regarding their models and their ratings. My associate asked several questions. "What are the key drivers of your rating model?" They responded, FICO scores and home price appreciation (HPA) of low single digit (LSD) or mid single digit (MSD), as HPA has been for the past 50 years. My associate then asked, "What if HPA was flat for an extended period of time?" They responded that their model would start to break down. He then asked, "What if HPA were to decline 1% to 2% for an extended period of time?" They responded that their models would break down completely. He then asked, "With 2% depreciation, how far up the rating's scale would it harm?" They responded that it might go as high as the AA or AAA tranches.

It certainly appears that rating agencies did not fully appreciate the fragility of their estimates nor the possible effects of modest errors in assumptions about default correlations and probabilities in their credit ratings. But this lack of understanding was apparently shared by the regulators who tied bank capital requirements to ratings, as well as by the investors who outsourced their due diligence to rating agencies without sufficient consideration of whether credit ratings meant the same thing for structured finance as they had for single-name securities. In particular, none of the key parties seemed to recognize that small errors in rating individual securities, errors that would have no material effect in the single-name market, are significantly magnified in the tranches of a collateralized debt obligation structure, and can be further magnified when CDO^2 are created from the original collateralized debt obligations, as was common in the mortgage-backed securitizations.

There is also some evidence that perverse incentives induced questionable behavior on the part of market participants. One concern is over the possibility of conflicts of interest that may arise because the issuer, rather than the investor, pays for the rating. Mason and Rosner (2007) argue that the process and complexity of creating structured finance products requires rating agencies essentially to "become part of the underwriting team" rather than act as agents for outside investors. On the other side, the Committee on the Global Financial System from the Bank of International Settlements (2005, p. 26) investigated this concern and concluded that it was no more severe for structured finance products than for single-name credit products, arguing that reputation was a strong force against bad behavior in both markets: "In fact, there appear to be no fundamental differences in the rating processes for structured finance products and traditional bonds. The potential conflicts of interest arising in structured finance are thus unlikely to be materially different from those in the traditional segments of the agencies' business." Looking at the Bank of International Settlements (2005) report several years later, it offers an example of how a variety of important market participants viewed structured finance products and traditional bonds to be highly similar. It also articulates a widely-held view that market forces would solve potential problems. This confusion over the nature of structured products combined with a belief and reliance on market efficiency proved a potent combination.

The U.S. Securities and Exchange Commission (2008) recently summarized its findings from an investigation of several credit rating agencies. It found much that could be improved in the rating process and that analysts and managers generally understood how their actions affected profits and could be in conflict with the goal of accurate credit risk assessment (p. 12):

> For example, in one exchange of internal communications between two analysts at one rating agency, the analysts were concerned about whether they should be rating a particular deal. One analyst expressed concern that her firm's model did not capture "half" of the deal's risk and that "it could be structured by cows and we would rate it" (Email no. 1: Analytical Staff to Analytical Staff, Apr. 5, 2007, 1:13 PM).

> In another email, an analytical manager in the same rating agency's CDO group wrote to a senior analytical manager that the rating agencies continue to create an "even bigger monster—the CDO market. Let's hope we are all wealthy and retired by the time this house of cards falters. ;o)" (Email no. 2: Analytical Manager to Senior Analytical Manager, Dec. 15, 2006, 8:31 PM).

The investment banks played a dual role of investors and dealers in the structured finance market. The business offered enormous short-run payoffs, which seemed too compelling to ignore even if value-destroying in the long-run. The banks were generally eager to keep the structured finance business going even as underwriting standards fell. The combination of low capital requirements imposed on AAA-rated assets and a commonly held perception that they were "safe," allowed banks to hold on to any senior tranches that were not sold to investors. But when the structured finance market collapsed in late 2007, the investment banks found themselves holding hundreds of billions of dollars of low-quality asset pools, many of which consisted of leveraged buy-out loans, subprime mortgages, and bonds from collateralized debt obligations in process—that is, where the tranches had not yet been sold to other investors.[9]

There is some evidence that Wall Street executives realized it would end one day, but in the meantime, they had little incentive to move to the sidelines. In July 2007, the then-CEO of Citigroup, Chuck Prince, acknowledged that the cheap credit-fueled buy-out boom would eventually end, but that in the meantime, his firm would continue to participate in structured finance activities (as reported in Nakamoto and Wighton, 2007): "When the music stops, in terms of liquidity, things will get complicated. As long as the music is playing, you've got to get up and dance. We're still dancing."

Finally, the minimum capital requirements for banks set forth in Basel I and II may have played an important role in the evolution of the structured finance market. Under these guidelines, banks holding AAA-rated securities were required to hold only half as much capital as was required to support other investment-grade securities. As a result of this built-in demand by banks for AAA-rated securities, a small yield advantage in AAA-rated structured finance securities may have led to a large increase in the demanded quantity. As discussed in the previous section, the structured finance machinery enabled the creation of AAA-rated securities that had a yield advantage over single-name AAA-rated securities, but only by filling these securities with systematic risks or by rating them incorrectly.

[9] For a detailed study of the market for collateralized loan obligations, see Benmelech and Dlugosz (2008).

Implications and Conclusions

During the credit crunch from late 2007 and into 2008, the buyers of highly rated structured finance products largely stopped buying. The initial cause for this change was that subprime-related securities were experiencing large losses, which created concerns about structured finance products more generally. Some practitioners believe that the credit crunch of 2007 and 2008 will work itself out, as such episodes tend to do, and the market for structured credit will return as before. We hold the more skeptical view that the market for structured credit appears to have serious structural problems that may be difficult to overcome.

As we have explained, these claims are highly sensitive to the assumptions of 1) default probability and recovery value, 2) correlation of defaults, and 3) the relation between payoffs and the economic states that investors care about most. Beginning in late 2007 and continuing well into 2008, it became increasingly clear to investors in highly-rated structured products that each of these three key assumptions were systematically biased against them. These investors are now reluctant to invest in securities that they do not fully understand.

The ability to create large quantities of AAA-rated securities from a given pool of underlying assets is likely to be forever diminished, as the rating process evolves to better account for parameter and model uncertainty. The key is recognizing that small errors that would not be costly in the single-name market, are significantly magnified by the collateralized debt obligation structure, and can be further magnified when collateralized debt obligations are created from the tranches of other collateralized debt obligations, as was common in mortgage-backed securitizations. The good news is that this mistake can be fixed. For example, a Bayesian approach that explicitly acknowledges that parameters are uncertain would go a long way towards solving this problem. Of course, adopting a Bayesian perspective on parameter uncertainty will necessarily mean far less AAA-rated securities can be issued and therefore fewer opportunities to offer investors attractive yields.

Additionally, investors need to recognize the fundamental difference between single-name and structured securities when it comes to exposure to systematic risk. Unlike traditional corporate bonds, whose fortunes are primarily driven by firm-specific considerations, the performance of securities created by tranching large asset pools is strongly affected by the performance of the economy as a whole. In particular, senior structured finance claims have the features of economic catastrophe bonds, in that they are designed to default only in the event of extreme economic duress. Because credit ratings are silent regarding the state of the world in which default is likely to happen, they do not capture this exposure to systematic risks. The lack of consideration for these types of exposures reduces the usefulness of ratings, no matter how precise they are made to be.

■ *We thank Eli Cohen, Lauren Cohen, Robin Greenwood, Jim Hines, Chris Malloy, Eric Del Monaco, André Perold, Hyun Shin, Andrei Shleifer, Jeremy Stein, and Timothy Taylor for helpful comments and discussions.*

References

Altman, Edward. 2006. "Default Recovery Rates and LGD in Credit Risk Modeling and Practice: An Updated Review of the Literature and Empirical Evidence." http://pages.stern.nyu.edu/~ealtman/UpdatedReviewofLiterature.pdf.

Altman, Edward, and Jeffrey Swanson. 2007. "The Investment Performance and Market Size of Defaulted Bonds and Bank Loans." http://pages.stern.nyu.edu/~ealtman/2006%20InvestPerf.pdf.

Ashcraft, Adam, and Til Schuermann. 2008. "Understanding the Securitization of Subprime Mortgage Credit." Federal Reserve Bank of New York Staff Report 318. http://www.newyorkfed.org/research/staff_reports/sr318.pdf.

Bank of International Settlements. 2005. "The Role of Ratings in Structured Finance: Issues and Implications." Committee on the Global Financial System Publications no. 23, January. Available at: http://www.bis.org/publ/cgfs23.htm.

Bank of International Settlements, Basel Committee on Banking Supervision, The Joint Forum. 2008. "Credit Risk Transfer—Developments from 2005—2007." July. Available at: http://www.bis.org/publ/joint21.htm.

Beales, Richard, Saskia Scholtes, and Gillian Tett. 2007. "Failing grades? Why Regulators Fear Credit Ratings May Be Out of Their Depth." *Financial Times,* May 17.

Benford, James, and Erlend Nier. 2007. "Monitoring Cyclicality of Basel II Capital Requirements." Bank of England, Financial Stability Paper No. 3.

Benmelech, Effraim, and Jennifer Dlugosz. 2008. "The Alchemy of CDO Credit Ratings." http://www.carnegie-rochester.rochester.edu/nov08-pdfs/benmelech_dlugosz.pdf.

Bhanota, Karan, and Antonio S. Mello. 2006. "Should Corporate Debt include a Rating Trigger?" *Journal of Financial Economics,* 79(1): 69–98.

Bharath, Sreedhar, and Tyler Shumway. 2007. "Forecasting Default with the Merton Distance to Default Model." *Review of Financial Studies,* 21(3): 1339–69.

Coval, Joshua, Jakub Jurek, and Erik Stafford. Forthcoming. "Economic Catastrophe Bonds." *American Economic Review.*

Craig, Susanne, Randall Smith, and Serena Ng. 2008. "Merrill Aims to Raise Billions More: Firm Dumps Mortgage Assets as Crisis Drags On; Another Big Write-Down." *Wall Street Journal,* July 29.

Derivative Fitch. 2006. "Global Criteria for Collateralized Debt Obligations."

Dichev, Ilia D., and Joseph D. Piotroski. 2001. "The Long-Run Stock Returns Following Bond Rating Changes." *Journal of Finance,* 56(1): 173–203.

Ederington, Louis H., and Jeremy Goh. 1998. "Bond Rating Agencies and Stock Analysts: Who Knows What When?" *The Journal of Finance and Quantitative Analysis,* 33(4): 569–85.

Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics,* 33(1): 3–56.

Federal Deposit Insurance Corporation. 2006. "FDIC Outlook: Breaking New Ground in Mortgage Lending." http://www.fdic.gov/bank/analytical/regional/ro20062q/na/2006_summer04.html.

Financial Week. 2008. "S&P President: Structured Finance Business in the Tank." June 10. http://www.financialweek.com/apps/pbcs.dll/article?AID=/20080610/REG/660111773/-1/FWDailyAlert01.

Fitch Ratings. 2007. "Inside the Ratings: What Credit Ratings Mean." http://www.bankwatchratings.com/descargas/ratings.pdf.

Hand, John R., Robert W. Holthausen, and Richard W. Leftwich. 1992. "The Effect of Bond Rating Agency Announcements on Bond and Stock Prices." *Journal of Finance,* 47(2): 733–52.

Heitfield, Erik. 2008. "Parameter Uncertainty and the Credit Risk of Collateralized Debt Obligations." Available at SSRN: http://ssrn.com/abstract=1190362.

Hite, Gailen, and Arthur Warga. 1997. "The Effect of Bond-Rating Changes on Bond Price Performance." *Financial Analysts Journal,* 53(3): 35–51.

Holthausen, Robert W., and Richard W. Leftwich. 1986. "The Effect of Bond Rating Changes on Common Stock Prices." *Journal of Financial Economics,* 17(1): 57–89.

Hu, Jian. 2007. "Assessing the Credit Risk of CDOs Backed by Structured Finance Securities: Rating Analysts' Challenges and Solutions." *Journal of Structured Finance,* Fall.

Hull, John, Mirela Predescu, and Alan White. 2004. "The Relationship Between Credit Default Swap Spreads, Bond Yields, and Credit Rating Announcements." *Journal of Banking and Finance,* 28(11): 2789–2811.

Jones, Sam, Gillian Tett, and Paul J. Davies. 2008. "Moody's Error Gave Top Ratings to Debt Products." *Financial Times,* May 20.

Jorion, Philippe, and Gaiyan Zhang. 2007. "Good and Bad Credit Contagion: Evidence from Credit Default Swaps." *Journal of Financial Economics,* 84(3): 860–83.

Katz, Steven. 1974. "The Price Adjustment Process of Bonds to Rating Reclassifications: A Test of Bond Market Efficiency." *Journal of Finance,* 29(2): 551–59.

Mason, Joseph, and Joshua Rosner. 2007. "Where Did the Risk Go? How Misapplied Bond Ratings Cause Mortgage Backed Securities and Collateralized Debt Obligation Market Disruptions." Available at SSRN: http://ssrn.com/abstract=1027475.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance,* 29(2): 449–70.

Moody's Investors Service. 2004. "Moody's Revisits Its Assumptions Regarding Corporate Default (and Asset) Correlations for CDOs."

Nakamoto, Michiyo, and David Wighton. 2007. "Citigroup Chief Stays Bullish on Buyouts." *Financial Times,* July 9.

Organisation for Economic Cooperation and Development. 2007. "Survey of Investment Regulations of Pension Funds." http://www.oecd.org/dataoecd/56/7/38969997.pdf.

Rodriguez, Robert. 2007. "Absence of Fear." A speech given at the CFA Society of Chicago, June 28. http://www.fpafunds.com/news_070703_absense_of_fear.asp.

Scholtes, Saskia, and Richard Beales. 2007. "Top Rating Proving Crucial to Structured Finance Sector." *Financial Times,* May 17.

Schonbucher, Philipp J. 2003. "Factor Models: Portfolio Credit Risks When Defaults Are Correlated." *Journal of Risk Finance,* 3(1): 45–56.

Securities Industry and Financial Markets Association. 2008. "Recommendations of the Securities Industry and Financial Markets Association Credit Rating Agency Task Force." http://www.sifma.org/capital_markets/docs/SIFMA-CRA-Recommendations.pdf.

Tarashev, Nikola, and Haibin Zhu. 2007. "Modelling and Calibration Errors in Measures of Portfolio Credit Risk." Bank of International Settlements Working Paper 230.

Tomlinson, Richard, and David Evans. 2007. "Rating Charade." *Bloomberg Markets Magazine,* July. http://www.bloomberg.com/news/marketsmag/ratings.html.

U.S. Securities and Exchange Commission. 2008. "Summary Report of Issues Identified in the Commission Staff's Examination of Select Credit Rating Agencies." http://www.sec.gov/news/studies/2008/craexamination070808.pdf.

Vasicek, Oldrich. 2002. "The Distribution of Loan Portfolio Value." Working Paper, KMV Corporation. http://www.moodyskmv.com/conf04/pdf/papers/dist_loan_port_val.pdf. (Reprinted in *RISK,* December, 15(12): 160–62.)

Wansley, James L., John Glascock, and Terrence M. Clauretie. 1992. "Institutional Bond Pricing and Information Arrival: The Case of Bond Rating Changes." *Journal of Business Finance and Accounting,* 19(5): 733–50.

This article has been cited by:

1. Gillian G.H. Garcia. 2009. Ignoring the lessons for effective prudential supervision, failed bank resolution and depositor protection. *Journal of Financial Regulation and Compliance* **17**:3, 186-209. [CrossRef]

CAS Exam 9 Study Kit

“Value at Risk: Uses and Abuses”

by C.L. Culp, M.H. Miller, and A.M.P Neves

VALUE AT RISK: USES AND ABUSES

by Christopher L. Culp,
CP Risk Management LLC,
Merton H. Miller,
University of Chicago, and
Andrea M. P. Neves,
*CP Risk Management LLC**

Value at risk ("VAR") is now viewed by many as indispensable ammunition in any serious corporate risk manager's arsenal. VAR is a method of measuring the financial risk of an asset, portfolio, or exposure over some specified period of time. Its attraction stems from its ease of interpretation as a summary measure of risk and consistent treatment of risk across different financial instruments and business activities. VAR is often used as an approximation of the "maximum reasonable loss" a company can expect to realize from all its financial exposures.

VAR has received widespread accolades from industry and regulators alike.[1] Numerous organizations have found that the practical uses and benefits of VAR make it a valuable decision support tool in a comprehensive risk management process. Despite its many uses, however, VAR—like any statistical aggregate—is subject to the risk of misinterpretation and misapplication. Indeed, most problems with VAR seem to arise from what a firm *does* with a VAR measure rather than from the actual computation of the number.

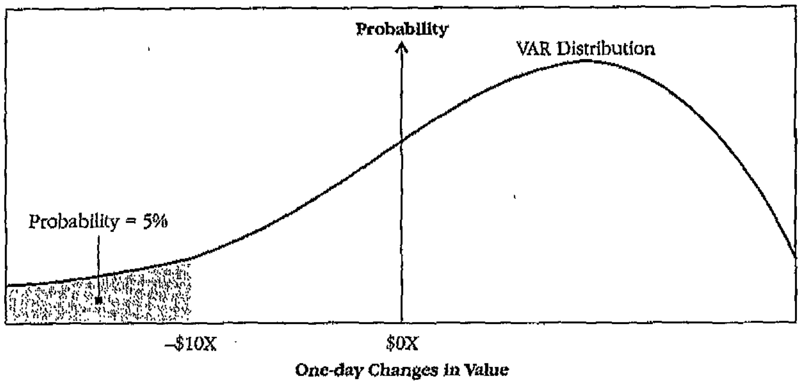
Why a company manages risk affects *how* a company should manage—and, hence, should measure—its risk.[2] In that connection, we examine the four "great derivatives disasters" of 1993-1995—Procter & Gamble, Barings, Orange County, and Metallgesellschaft—and evaluate how *ex ante* VAR measurements likely would have affected those situations. We conclude that VAR would have been of only limited value in averting those disasters and, indeed, actually might have been *misleading* in some of them.

*The authors thank Kamaryn Tanner for her previous work with us on this subject.

1. *See, for example,* Global Derivatives Study Group, *Derivatives: Practices and Principles* (Washington, D.C.: July 1993), and Board of Governors of the Federal Reserve System, *SR Letter 93-69* (1993). Most recently, the Securities and Exchange Commission began to require risk disclosures by all public companies. One approved format for these mandatory financial risk disclosures is VAR. For a critical assessment of the SEC's risk disclosure rule, *see* Merton H. Miller and Christopher L. Culp, "The SEC's Costly Disclosure Rules," *Wall Street Journal* (June 22, 1996).

2. This presupposes, of course, that "risk management" is consistent with value-maximizing behavior by the firm. For the purpose of this paper, we do not consider whether firms *should be* managing their risks. For a discussion of that issue, *see* Christopher L. Culp and Merton H. Miller, "Hedging in the Theory of Corporate Finance: A Reply to Our Critics," *Journal of Applied Corporate Finance*, Vol. 8, No. 1 (Spring 1995):121-127, and Rene M. Stulz, "Rethinking Risk Management," *Journal of Applied Corporate Finance*, Vol. 9, No. 3 (Fall 1996):8-24.

FIGURE 1



WHAT IS VAR?

Value at risk is a summary statistic that quantifies the exposure of an asset or portfolio to market risk, or the risk that a position declines in value with adverse market price changes.[3] Measuring risk using VAR allows managers to make statements like the following: "We do not expect losses to exceed $1 million on more than 1 out of the next 20 days."[4]

To arrive at a VAR measure for a given portfolio, a firm must generate a probability distribution of possible changes in the value of some portfolio over a specific time or "risk horizon"—*e.g.*, one day.[5] The value at risk of the portfolio is the dollar loss corresponding to some pre-defined probability level—usually 5% or less—as defined by the left-hand tail of the distribution. Alternatively, VAR is the dollar loss that is expected to occur no more than 5% of the time over the defined risk horizon. Figure 1, for example, depicts a one-day VAR of $10X at the 5% probability level.

The Development of VAR

VAR emerged first in the trading community.[6] The original goal of VAR was to systematize the measurement of an active trading firm's risk exposures across its dealing portfolios. Before VAR, most commercial trading houses measured and controlled risk on a desk-by-desk basis with little attention to firm-wide exposures. VAR made it possible for dealers to use risk measures that could be compared and aggregated across trading areas as a means of monitoring and limiting their consolidated financial risks.

VAR received its first public endorsement in July 1993, when a group representing the swap dealer community recommended the adoption of VAR by all active dealers.[7] In that report, the Global Derivatives Study Group of The Group of Thirty urged dealers to "use a *consistent measure* to calculate daily the market risk of their derivatives positions and compare it to market risk limits. Market risk is best measured as 'value at risk' using *probability analysis* based upon a common confidence interval (*e.g.*, two standard deviations) and *time horizon* (*e.g.*, a one-day exposure). [emphasis added]"[8]

The italicized phrases in The Group of Thirty recommendation draw attention to several specific features of VAR that account for its widespread popularity among trading firms. One feature of VAR is its *consistent* measurement of

3. More recently, VAR has been suggested as a framework for measuring credit risk, as well. To keep our discussion focused, we examine only the applications of VAR to market risk measurement.

4. For a general description of VAR, *see* Philippe Jorion, *Value at Risk* (Chicago: Irwin Professional Publishing, 1997).

5. The risk horizon is chosen exogenously by the firm engaging in the VAR calculation.

6. An early precursor of VAR was SPANTM—Standard Portfolio Analysis of Risk—developed by the Chicago Mercantile Exchange for setting futures margins. Now widely used by virtually all futures exchanges, SPAN is a non-parametric, simulation-based "worst case" measure of risk. As will be seen, VAR, by contrast, rests on well-defined probability distributions.

7. This was followed quickly by a similar endorsement from the International Swaps and Derivatives Association. *See* Jorion, cited previously.

8. Global Derivatives Study Group, cited previously.

financial risk. By expressing risk using a "possible dollar loss" metric, VAR makes possible direct comparisons of risk across different business lines and distinct financial products such as interest rate and currency swaps.

In addition to consistency, VAR also is *probability-based.* With whatever degree of confidence a firm wants to specify, VAR enables the firm to associate a specific loss with that level of confidence. Consequently, VAR measures can be interpreted as forward-looking approximations of potential market risk.

A third feature of VAR is its reliance on a *common time horizon* called the risk horizon. A one-day risk horizon at, say, the 5% probability level tells the firm, strictly speaking, that it can expect to lose no more than, say, $10X on the next day with 95% confidence. Firms often go on to assume that the 5% confidence level means they stand to lose more than $10X on no more than five days out of 100, an inference that is true only if strong assumptions are made about the stability of the underlying probability distribution.[9]

The choice of this risk horizon is based on various motivating factors. These may include the timing of employee performance evaluations, key decision-making events (*e.g.*, asset purchases), major reporting events (*e.g.*, board meetings and required disclosures), regulatory examinations, tax assessments, external quality assessments, and the like.

Implementing VAR

To estimate the value at risk of a portfolio, possible future values of that portfolio must be generated, yielding a distribution—called the "VAR distribution"—like that we saw in Figure 1. Once the VAR distribution is created for the chosen risk horizon, the VAR itself is just a number on the curve—*viz.*, the change in the value of the portfolio leaving the specified amount of probability in the left-hand tail.

Creating a VAR distribution for a particular portfolio and a given risk horizon can be viewed as a two-step process.[10] In the first step, the price or return distributions for each individual security or asset in the portfolio are generated. These distributions represent possible value changes in all the component assets over the risk horizon.[11] Next, the individual distributions somehow must be aggregated into a portfolio distribution using appropriate measures of correlation.[12] The resulting portfolio distribution then serves as the basis for the VAR summary measure.

An important assumption in almost all VAR calculations is that the portfolio whose risk is being evaluated *does not change* over the risk horizon. This assumption of no turnover was not a major issue when VAR first arrived on the scene at derivatives dealers. They were focused on one- or two-day—sometimes *intra*-day—risk horizons and thus found VAR both easy to implement and relatively realistic. But when it comes to generalizing VAR to a longer time horizon, the assumption of no portfolio changes becomes problematic. What does it mean, after all, to evaluate the *one-year* VAR of a portfolio using only the portfolio's contents *today* if the turnover in the portfolio is 20-30% per day?

Methods for generating both the individual asset risk distributions and the portfolio risk distribution range from the simplistic to the indecipherably complex. Because our goal in this paper is not to evaluate all these mechanical methods of VAR measurement, readers are referred elsewhere for explanations of the nuts and bolts of VAR computation.[13] Several common methods of VAR calculation are summarized in the Appendix.

9. This interpretation assumes that asset price changes are what the technicians call "iid," independently and identically distributed—*i.e.*, that price changes *are drawn from essentially the same distribution every day.*

10. In practice, VAR is not often implemented in a clean two-step manner, but discussing it in this way simplifies our discussion—without any loss of generality.

11. Especially with instruments whose payoffs are non-linear, a better approach is to generate distributions for the underlying "risk factors" that affect an asset rather than focus on the changes in the values of the assets themselves. To *generate the value change distribution of an option on a stock, for example, one* might first generate changes in the stock price and its volatility and *then* compute associated option price changes rather than generating option price changes "directly." For a discussion, *see* Michael S. Gamze and Ronald S. Rolighed, "VAR: What's Wrong With This Picture?," unpublished manuscript, Federal Reserve Bank of Chicago (1997).

12. If a risk manager is interested in the risk of a particular financial instrument, the appropriate risk measure to analyze is *not* the VAR of that instrument. Portfolio *effects still must be considered. The relevant measure of risk is the marginal risk* of that instrument in the portfolio being evaluated. *See* Mark Garman, "Improving on VAR," *Risk*, Vol. 9, No. 5 (1996):61-63.

13. *See, for example,* Jorion, cited previously, and Rod A. Beckström and Alyce R. Campbell, "Value-at-Risk (VAR): Theoretical Foundations," in *An Introduction to VAR*, Rod Beckström and Alyce Campbell, eds. (Palo Alto, Ca.: CAATS Software, Inc., 1995), *and James V. Jordan and Robert J. Mackay, "Assessing Value at Risk* for Equity Portfolios: Implementing Alternative Techniques," in *Derivatives Handbook*, Robert J. Schwartz and Clifford W. Smith, Jr., eds. (New York: John Wiley & Sons, Inc., 1997).

VAR made it possible for dealers to use risk measures that could be compared and aggregated across trading areas as a means of monitoring and limiting their consolidated financial risks.

Uses of VAR

The purpose of any risk measurement system and summary risk statistic is to facilitate risk reporting and control decisions. Accordingly, dealers quickly began to rely on VAR measures in their broader risk management activities. The simplicity of VAR measurement greatly facilitated dealers' reporting of risks to senior managers and directors. The popularity of VAR owes much to Dennis Weatherstone, former chairman of JP Morgan & Co., Inc., who demanded to know the total market risk exposure of JP Morgan at 4:15pm every day. Weatherstone's request was met with a daily VAR report.

VAR also proved useful in dealers' risk *control* efforts.[14] Commercial banks, for example, used VAR measures to quantify current trading exposures and compare them to established counterparty risk limits. In addition, VAR provided traders with information useful in formulating hedging policies and evaluating the effects of particular transactions on net portfolio risk. For managers, VAR became popular as a means of analyzing the performance of traders for compensation purposes and for allocating reserves or capital across business lines on a risk-adjusted basis.

Uses of VAR by Non-Dealers. Since its original development as a risk management tool for active trading firms, VAR has spread outside the dealer community. VAR now is used regularly by non-financial corporations, pension plans and mutual funds, clearing organizations, brokers and futures commission merchants, and insurers. These organizations find VAR just as useful as trading firms, albeit for different reasons.

Some benefits of VAR for non-dealers relate more to the exposure monitoring facilitated by VAR measurement than to the risk measurement task itself. For example, a pension plan with funds managed by external investment advisors may use VAR for policing its external managers. Similarly, brokers and account merchants can use VAR to assess collateral requirements for customers.

VAR AND CORPORATE RISK MANAGEMENT OBJECTIVES

Firms managing risks may be either *value risk managers* or *cash flow risk managers*.[15] A value risk manager is concerned with the firm's total value at a particular point in time. This concern may arise from a desire to avoid bankruptcy, mitigate problems associated with informational asymmetries, or reduce expected tax liabilities.[16] A cash flow risk manager, by contrast, uses risk management to reduce cash flow volatility and thereby increase debt capacity.[17] Value risk managers thus typically manage the risks of a *stock of assets*, whereas cash flow risk managers manage the risks of a *flow of funds*. A risk measure that is appropriate for one type of firm may not be appropriate for others.

Value Risk Managers and VAR-Based Risk Controls

As the term *value* at risk implies, organizations for which VAR is best suited are those for which *value* risk management is the goal. VAR, after all, is intended to summarize the risk of a stock of assets over a particular risk horizon. Those likely to realize the most benefits from VAR thus include clearing-houses, securities settlement agents,[18] and swap dealers. These organizations have in common a concern with the value of their exposures over a well-defined period of time *and* a wish to limit and control those exposures. In addition, the relatively short risk horizons of these enterprises imply that VAR measurement can be accomplished reliably and with minimal concern about changing portfolio composition over the risk horizon.

Many value risk managers have risks arising mainly from *agency* transactions. Organizations like financial clearing-houses, for example, are exposed to risk arising from intermediation services rather than the risks of proprietary position taking. VAR can assist such firms in monitoring their customer credit exposures, in

14. *See* Rod A. Beckström and Alyce R. Campbell, "The Future of Firm-Wide Risk Management," in *An Introduction to VAR*, Rod Beckström and Alyce Campbell, eds. (Palo Alto, Ca.: CAATS Software, Inc., 1995).

15. For a general discussion of the traditional corporate motivations for risk management, *see* David Fite and Paul Pfleider, "Should Firms Use Derivatives to Manage Risk?" in *Risk Management: Problems & Solutions*, William H. Beaver and George Parker, eds. (New York: McGraw-Hill, Inc., 1995).

16. *See, for example*, Clifford Smith and Rene Stulz, "The Determinants of Firms' Hedging Policies," *Journal of Financial and Quantitative Analysis*, Vol. 20 (1985):391-405.

17. *See, for example*, Kenneth Froot, David Scharfstein, and Jeremy Stein, "Risk Management: Coordinating Corporate Investment and Financing Policies," *Journal of Finance*, Vol. 48 (1993):1629-1658.

18. *See* Christopher L. Culp and Andrea M.P. Neves, "Risk Management by Securities Settlement Agents," *Journal of Applied Corporate Finance*, Vol. 10, No. 3 (Fall 1997):96-103.

setting position and exposure limits, and in determining and enforcing margin and collateral requirements.

Total vs. Selective Risk Management

Most financial distress-driven explanations of corporate risk management, whether value or cash flow risk, center on a firm's *total risk.*[19] If so, such firms should be indifferent to the *composition* of their total risks. *Any* risk thus is a candidate for risk reduction.

Selective risk managers, by contrast, deliberately choose to manage some risks and not others. Specifically, they seek to manage their exposures to risks in which they have no comparative informational advantage—for the usual financial ruin reasons—while actively exposing themselves, at least to a point, to risks in which they *do* have perceived superior information.[20]

For firms managing total risk, the principal benefit of VAR is facilitating explicit risk control decisions, such as setting and enforcing exposure limits. For firms that selectively manage risk, by contrast, VAR is useful largely for diagnostic monitoring *or* for controlling risk in areas where the firm perceives no comparative informational advantage. An airline, for example, might find VAR helpful in assessing its exposure to jet fuel prices; but for the airline to use VAR to analyze the risk that seats on its aircraft are not all sold makes little sense.

Consider also a hedge fund manager who invests in foreign equity because the risk/return profile of that asset class is desirable. To avoid exposure to the exchange rate risk, the fund could engage an "overlay manager" to hedge the currency risk of the position. Using VAR on the *whole position* lumps together two separate and distinct sources of risk—the currency risk and the foreign equity price risk. And *reporting* that total VAR without a corresponding expected return could have disastrous consequences. Using VAR to ensure that the currency hedge is accomplishing its intended aims, by contrast, might be perfectly legitimate.

VAR AND THE GREAT DERIVATIVES DISASTERS

Despite its many benefits to certain firms, VAR is not a panacea. Even when VAR is calculated appropriately, VAR *in isolation* will do little to keep a firm's risk exposures in line with the firm's chosen risk tolerances. Without a well-developed risk management infrastructure—policies and procedures, systems, and well-defined senior management responsibilities—VAR will deliver little, if any, benefits. In addition, VAR may not always help a firm accomplish its particular risk management objectives, as we shall see.

To illustrate some of the pitfalls of using VAR, we examine the four "great derivatives disasters" of 1993-1995: Procter & Gamble, Orange County, Barings, and Metallgesellschaft.[21] Proponents of VAR often claim that many of these disasters would have been averted had VAR measurement systems been in place. We think otherwise.[22]

Procter & Gamble

During 1993, Procter & Gamble ("P&G") undertook derivatives transactions with Bankers Trust that resulted in over $150 million in losses.[23] Those losses traced essentially to P&G's writing of a put option on interest rates to Bankers Trust. Writers of put options suffer losses, of course, whenever the underlying security declines in price, which in this instance meant whenever interest rates rose. And rise they did in the summer and autumn of 1993.

The put option actually was only one component of the whole deal. The deal, with a notional principal of $200 million, was a fixed-for-floating rate swap in which Bankers Trust offered P&G 10 years of floating-rate financing at 75 basis points below the commercial paper rate in exchange for the put and fixed interest payments of 5.3% annually. That huge financing advantage of 75 basis points apparently was too much for P&G's treasurer to resist, particularly because the put was well out-of-the-money

19. *See, for example,* Smith and Stulz, cited previously, and Froot, Scharfstein, and Stein, cited previously.

20. *See* Culp and Miller (Spring 1995), cited previously, and Stulz, cited previously.

21. In truth, Procter & Gamble was the only one of these disasters actually *caused* by derivatives. *See* Merton H. Miller, "The Great Derivatives Disasters: What Really Went Wrong and How to Keep It from Happening to You," speech presented to JP Morgan & Co. (Frankfurt, June 24, 1997) and Chapter Two in Merton H. Miller, *Merton Miller on Derivatives* (New York: John Wiley & Sons, Inc., 1997).

22. The details of all these cases are complex. We thus refer readers elsewhere for discussions of the actual events that took place and limit our discussion here only to basic background. *See, for example,* Stephen Figlewski, "How to Lose Money in Derivatives," *Journal of Derivatives*, Vol. 2, No. 2 (Winter 1994):75-82.

23. *See, for example,* Figlweski, cited previously, and Michael S. Gamze and Karen McCann, "A Simplified Approach to Valuing an Option on a Leveraged Spread: The Bankers Trust, Procter & Gamble Example," *Derivatives Quarterly*, Vol. 1, No. 4 (Summer 1995):44-53.

For firms that selectively manage risk, VAR is useful for controlling risk only in areas where the firm perceives no comparative informational advantage. An airline might find VAR helpful in assessing its exposure to jet fuel prices; but to use VAR to analyze the risk that seats on its aircraft are not all sold makes little sense.

when the deal was struck in May 1993. But the low financing rate, of course, was just premium collected for writing the put. When the put went in-the-money for Bankers Trust, what once seemed like a good deal to P&G ended up costing millions of dollars.

VAR would have helped P&G, *if* P&G also had in place an adequate risk management infrastructure—which apparently it did not. Most obviously, senior managers at P&G would have been unlikely to have approved the original swap deal if its exposure had been subject to a VAR calculation. But that presupposes a lot.

Although VAR would have helped P&G's senior management measure its exposure to the speculative punt by its treasurer, much more would have been needed to stop the treasurer from taking the interest rate bet. The first requirement would have been a system for measuring the risk of the swaps *on a transactional basis.* But VAR was never intended for use on single transactions.[24] On the contrary, the whole appeal of the concept initially was its capacity to aggregate risk *across* transactions and exposures. To examine the risk of an individual transaction, the *change* in portfolio VAR that would occur with the addition of that new transaction should be analyzed. But that still requires first calculating the total VAR.[25] So, for P&G to have looked at the risk of its swaps in a VAR context, its entire treasury area would have needed a VAR measurement capability.

Implementing VAR for P&G's entire treasury function might *seem* to have been a good idea *anyway.* Why not, after all, perform a comprehensive VAR analysis on the whole treasury area and get transactional VAR assessment capabilities as an added bonus? For some firms, that *is* a good idea. But for other firms, it is not. Many non-financial corporations like P&G, after all, typically undertake risk management in their corporate treasury functions for *cash flow* management reasons.[26] VAR is a *value* risk measure, not a cash flow risk measure. For P&G to examine value at risk for its *whole* treasury operation, therefore, presumes that P&G was a *value* risk manager, and that may not have been the case. Even had VAR been in place at P&G, moreover, the assumption that P&G's senior managers would have been *monitoring* and *controlling* the VARs of individual swap transactions is not a foregone conclusion.

Barings

Barings PLC failed in February 1995 when rogue trader Nick Leeson's bets on the Japanese stock market went sour and turned into over $1 billion in trading losses.[27,28] To be sure, VAR would have led Barings senior management to shut down Leeson's trading operation in time to save the firm—*if they knew about it.* If P&G's sin was a lack of internal management and control over its treasurer, then Barings was guilty of an even more cardinal sin. The top officers of Barings lost control over the trading operation *not* because no VAR measurement system was in place, but because they let the same individual making the trades also serve as the recorder of those trades—violating one of the most elementary principles of good management.

The more interesting question emerging from Barings is why top management seems to have taken so long to recognize that a rogue trader was at work. For that purpose, a fully functioning VAR system would certainly have helped. Increasingly, companies in the financial risk-taking business use VAR as a monitoring tool for detecting unauthorized increases in positions.[29] Usually, this is intended for *customer* credit risk management by firms like futures commission merchants. In the case of Barings, however, such account monitoring would have enabled management to spot Leeson's run-up in positions in his so-called "arbitrage" and "error" accounts.

VAR measurements at Barings, on the other hand, would have been impossible to implement,

24. Recently, some have advocated that derivatives dealers should evaluate the VAR of specific transactions *from the perspective of their counterparties* in order to determine counterparty suitability. Without knowing the rest of the counterparty's risk exposures, however, the VAR estimate would be meaningless. Even with full knowledge of the counterparty's total portfolio, the VAR number still might be of no use in determining suitability for reasons to become clear later.

25. *See* Garman, cited previously.

26. *See, for example,* Judy C. Lewent and A. John Kearney, "Identifying, Measuring, and Hedging Currency Risk at Merck," *Journal of Applied Corporate Finance*, Vol. 2, No. 4 (Winter 1990):19-28, and Deana R. Nance, Clifford W. Smith, and Charles W. Smithson, "On the Determinants of Corporate Hedging," *Journal of Finance*, Vol. 48, No. 1 (1993):267-284.

27. *See, for example,* Hans R. Stoll, "Lost Barings: A Tale in Three Parts Concluding with a Lesson," *Journal of Derivatives*, Vol. 3, No. 1 (Fall 1995):109-115., and Anatoli Kuprianov, "Derivatives Debacles: Case Studies of Large Losses in Derivatives Markets," in *Derivatives Handbook: Risk Management and Control*, Robert J. Schwartz and Clifford W. Smith, Jr., eds. (New York: John Wiley & Sons, Inc., 1997).

28. Our reference to rogue traders is not intended to suggest, of course, that rogue traders are only found in connection with derivatives. Rogue traders have caused the banks of this world far more damage from failed real estate (and copper) deals than from derivatives.

29. *See* Christopher Culp, Kamaryn Tanner, and Ron Mensink, "Risks, Returns and Retirement," *Risk*, Vol. 10, No. 10 (October 1997):63-69.

given the deficiencies in the *overall* information technology ("IT") systems in place at the firm. At any point in time, Barings' top managers knew only what Leeson was telling them. If Barings' systems were incapable of reconciling the position build-up in Leeson's accounts with the huge wire transfers being made by London to support Leeson's trading in Singapore, no VAR measure would have included a complete picture of Leeson's positions. And without that, no warning flag would have been raised.

Orange County

The Orange County Investment Pool ("OCIP") filed bankruptcy in December 1994 after reporting a drop in its market value of $1.5 billion. For many years, Orange County maintained the OCIP as the equivalent of a money market fund for the benefit of school boards, road building authorities, and other local government bodies in its territory. These local agencies deposited their tax and other collections when they came in and paid for their own wage and other bills when the need arose. The Pool paid them interest on their deposits—handsomely, in fact. Between 1989-1994, the OCIP paid its depositors 400 basis points more than they would have earned on the corporate Pool maintained by the State of California—roughly $750 million over the period.[30]

Most of the OCIP's investments involved leveraged purchases of intermediate-term securities and structured notes financed with "reverse repos" and other short-term borrowings. Contrary to conventional wisdom, the Pool was making its profits *not* from "speculation on falling interest rates" but rather from an investment play on the *slope* of the yield curve.[31] When the Federal Reserve started to raise interest rates in 1994, the intermediate-term securities declined in value and OCIP's short-term borrowing costs rose.

Despite the widespread belief that the leverage policy led to the fund's insolvency and bankruptcy filing, Miller and Ross, after examining the OCIP's investment strategy, cash position, and net asset value at the time of the filing, have shown that the OCIP was *not* insolvent. Miller and Ross estimate that the $20 billion in total assets on deposit in the fund had a positive net worth of about $6 billion. Nor was the fund in an illiquid cash situation. OCIP had over $600 million of cash on hand and was generating further cash at a rate of more than $30 million a month.[32] Even the reported $1.5 billion "loss" would have been completely recovered within a year—a loss that was realized only because Orange County's bankruptcy lawyers forced the liquidation of the securities.[33]

Jorion has taken issue with Miller and Ross's analysis of OCIP, arguing that VAR would have called the OCIP investment program into question long before the $1.5 billion loss was incurred.[34] Using several different VAR calculation methods, Jorion concludes that OCIP's one-year VAR at the end of 1994 was about $1 billion at the 5% confidence level. Under the usual VAR interpretation, this would have told OCIP to expect a loss in excess of $1 billion in one out of the next 20 years.

Even assuming Jorion's VAR number is accurate, however, his interpretation of the VAR measure was unlikely to have been the OCIP's interpretation—at least not *ex ante* when it could have mattered. The VAR measure in isolation, after all, takes no account of the *upside* returns OCIP was receiving as compensation for that downside risk. Remember that OCIP was pursuing a very deliberate yield curve, net cost-of-carry strategy, designed to generate high expected cash returns. That strategy had risks, to be sure, but those risks seem to have been clear to OCIP treasurer Robert Citron—and, for that matter, to the people of Orange County who re-elected Citron treasurer in preference to an opposing candidate who was criticizing the investment strategy.[35]

Had Orange County been using VAR, however, it almost certainly *would* have terminated its investment program upon seeing the $1 billion risk

30. Miller, cited previously.

31. When the term structure is upward sloping, borrowing in short-term markets to leverage longer-term government securities generates positive cash carry. A surge in inflation or interest rates, of course, could reverse the term structure and turn the carry negative. That is the real risk the treasurer was taking. But it was not much of a risk. Since the days of Jimmy Carter in the late 1970s, the U.S. term structure has never been downward sloping and nobody in December 1994 thought it was likely to be so in the foreseeable future.

32. Merton H. Miller and David J. Ross, "The Orange County Bankruptcy and its Aftermath: Some New Evidence," *Journal of Derivatives*, Vol. 4, No. 4 (Summer 1997):51-60.

33. Readers may wonder why, then, Orange County did declare bankruptcy. That story is complicated, but a hint might be found in the payment of $50 million in special legal fees to the attorneys that sued Merrill Lynch for $1.5 billion for selling OCIP the securities that lost money. In short, *lots* of people gained from OCIP's bankruptcy, even though OCIP was not actually bankrupt. *See* Miller, cited previously, and Miller and Ross, cited previously.

34. Philippe Jorion, "Lessons From the Orange County Bankruptcy," *Journal of Derivatives*, Vol. 4, No. 4 (Summer 1997):61-66.

35. Miller and Ross, cited previously.

Especially for institutional investors, a major pitfall of VAR is to highlight large potential losses over long time horizons *without conveying any information about the corresponding expected return.*

estimate. The reason probably would *not* have been the actual informativeness of the VAR number, but rather the fear of a public outcry at the number. Imagine the public reaction if the OCIP announced one day that it expected to lose more than $1 billion over the next year in one time out of 20. But that reaction would have far less to do with the actual risk information conveyed by the VAR number than with the lack of any corresponding expected profits reported *with* the risk number. Just consider, after all, what the public reaction would have been if the OCIP publicly announced that it would *gain* more than $1 billion over the next year in one time out of 20![36]

This example highlights a major abuse of VAR—an abuse that has nothing to do with the meaning of the value at risk number but instead traces to the presentation of the information that number conveys. Especially for institutional investors, a major pitfall of VAR is to highlight large potential losses over long time horizons *without conveying any information about the corresponding expected return.* The lesson from Orange County to would-be VAR users thus is an important one—for organizations whose mission is *to take some risks,* VAR measures of risks are meaningful *only* when interpreted alongside estimates of corresponding potential *gains.*

Metallgesellschaft

MG Refining & Marketing, Inc. ("MGRM"), a U.S. subsidiary of Metallgesellschaft AG, reported $1.3 billion in losses by year-end 1993 from its oil trading activities. MGRM's oil derivatives were part of a marketing program under which it offered long-term customers firm price guarantees for up to 10 years on gasoline, heating oil, and diesel fuel purchased from MGRM.[37] The firm hedged its resulting exposure to spot price increases with short-term futures contracts to a considerable extent. After several consecutive months of *falling* prices in the autumn of 1993, however, MGRM's German parent reacted to the substantial margin calls on the losing futures positions by liquidating the hedge, thereby turning a paper loss into a very real one.[38]

Most of the arguments over MGRM—in press accounts, in the many law suits the case engendered, and in the academic literature—have focused on whether MGRM was "speculating" or "hedging." The answer, of course, is that like all other merchant firms, they were doing both. They were definitely speculating on the oil "basis"—inter-regional, inter-temporal, and inter-product *differences* in prices of crude, heating oil, and gasoline. That was the *business they were in.*[39] The firm had expertise and informational advantages far beyond those of its customers or of casual observers playing the oil futures market. What MGRM did not have, of course, was any special expertise about the level and direction of oil prices generally. Here, rather than take a corporate "view" on the direction of oil prices, like the misguided one the treasurer of P&G took on interest rates, MGRM chose to hedge its exposure to oil price *levels.*

Subsequent academic controversy surrounding the case has mainly been not whether MGRM was hedging, but whether they were *over*-hedging—whether the firm could have achieved the same degree of insulation from price level changes with a lower commitment from MGRM's ultimate owner-creditor Deutsche Bank.[40] The answer is that the day-to-day cash-flow volatility of the program *could* have been reduced by any number of cash flow variance-reducing hedge ratios.[41] But the cost of chopping off some cash drains when prices fell was that of losing the corresponding cash inflows when prices spiked up.[42]

Conceptually, of course, MGRM could have used VAR analysis to measure its possible financial risks. But why would they have wanted to do so? The combined marketing/hedging program, after all, was *hedged* against changes in the *level* of oil prices. The only significant risks to which MGRM's

36. Only for the purpose of this example, we obviously have assumed symmetry in the VAR distribution.

37. A detailed analysis of the program can be found in Christopher L. Culp and Merton H. Miller, "Metallgesellschaft and the Economics of Synthetic Storage," *Journal of Applied Corporate Finance,* Vol. 7, No. 4 (Winter 1995):62-76.

38. For an analysis of the losses incurred by MRGM—as well as why they were incurred—*see* Christopher L. Culp and Merton H. Miller, "Auditing the Auditors," *Risk,* Vol. 8, No. 4 (1995):36-39.

39. Culp and Miller (Winter 1995, Spring 1995), cited previously, explain this in detail.

40. *See* Franklin R. Edwards and Michael S. Canter, "The Collapse of Metallgesellschaft: Unhedgeable Risks, Poor Hedging Strategy, or Just Bad Luck?," *Journal of Applied Corporate Finance,* Vol. 8, No. 1 (Spring 1995):86-105.

41. *See, for example,* Froot, Scharfstein, and Stein, cited previously.

42. A number of other reasons also explain MGRM's reluctance to adopt anything smaller than a "one-for-one stack-and-roll" hedge. *See* Culp and Miller (Winter 1995, Spring 1995).

program was subject thus were basis and rollover risks—again, the risk that MGRM was *in the business of taking.*[43]

A much bigger problem at MGRM than the change in the *value* of its program was the large negative *cash flows* on the futures hedge that would have been offset by eventual gains in the future on the fixed-price customer contracts. Although MGRM's former management claims it had access to adequate funding from Deutsche Bank (the firm's leading creditor and stock holder), perhaps some benefit might have been achieved by more rigorous cash flow simulations. But even granting that, VAR would have told MGRM very little about its *cash flows* at risk. As we have already emphasized, VAR is a *value*-based risk measure.

For firms like MGRM engaged in *authorized* risk-taking—like Orange County and unlike Leeson/Barings—the primary benefit of VAR really is just as an internal "diagnostic monitoring" tool. To that end, estimating the VAR of MGRM's basis trading activities *would* have told senior managers and directors at its parent what the basis risks were that MGRM actually was taking. But remember, MGRM's parent appears to have been fully aware of the risks MGRM's traders were taking *even without a VAR number.* In that sense, even the monitoring benefits of VAR for a proprietary trading operation would not have changed MGRM's fate.[44]

ALTERNATIVES TO VAR

VAR certainly is not the *only* way a firm can systematically measure its financial risk. As noted, its appeal is mainly its conceptual simplicity and its consistency across financial products and activities. In cases where VAR may *not* be appropriate as a measure of risk, however, other alternatives *are* available.

Cash Flow Risk

Firms concerned *not* with the value of a stock of assets and liabilities over a specific time horizon but with the volatility of a *flow of funds* often are better off eschewing VAR altogether in favor of a measure of cash flow volatility. Possible cash requirements over a *sequence* of future dates, for example, can be simulated. The resulting distributions of cash flows then enable the firm to control its exposure to cash flow risk more directly.[45] Firms worried about cash flow risk for preserving or increasing their debt capacities thus might engage in hedging, whereas firms concerned purely about liquidity shortfalls might use such cash flow stress tests to arrange appropriate standby funding.

Abnormal Returns and Risk-Based Capital Allocation

Stulz suggests managing risk-taking activities using abnormal returns—*i.e.*, returns in excess of the risk free rate—as a measure of the expected profitability of certain activities. Selective risk management then can be accomplished by allocating capital on a risk-adjusted basis and limiting capital at risk accordingly. To measure the risk-adjusted capital allocation, he suggests using the cost of new equity issued to finance the particular activity.[46]

On the positive side, Stulz's suggestion does not penalize selective risk managers for exploiting perceived informational advantages, whereas VAR does. The problem with Stulz's idea, however, lies in any company's capacity actually to implement such a risk management process. More properly, the difficulty lies in the actual estimation of the firm's equity cost of capital. And in any event, under M&M proposition three, all sources of capital are equivalent on a risk-adjusted basis. The source of capital for financing a particular project thus should not affect the decision to undertake that project. Stulz's reliance on equity only is thus inappropriate.

Shortfall Risk

VAR need not be calculated by assuming variance is a complete measure of "risk," but in practice this often *is* how VAR is calculated. (*See* the Appendix.) This assumption can be problematic when

43. The claim that MGRM was in the business of trading the basis has been disputed by managers of MGRM's parent and creditors. Nevertheless, the marketing materials of MGRM—on which the parent firm signed off—suggests otherwise. *See* Culp and Miller (Spring 1995), cited previously.

44. Like P&G and Barings, what happened at MGRM was, in the end, a *management* failure rather than a *risk management* failure. For details on how management failed in the MGRM case, *see* Culp and Miller (Winter 1995, Spring 1995). For a redacted version of the story, *see* Christopher L. Culp and Merton H. Miller, "Blame Mismanagement, Not Speculation, for Metall's Woes," *Wall Street Journal Europe* (April 25, 1995).

45. *See* Stulz, cited previously.

46. *See* Stulz, cited previously.

Firms concerned not with the value of a stock of assets and liabilities over a specific time horizon but with the volatility of a *flow of funds* often are better off eschewing VAR altogether in favor of a measure of cash flow volatility.

measuring exposures in markets characterized by non-normal (i.e., non-Gaussian) distributions—*e.g.*, return distributions that are skewed or have fat tails. If so, as explained in the Appendix, a solution is to generate the VAR distribution in a manner that does *not* presuppose variance is an adequate measure of risk. Alternatively, other summary risk measures can be calculated.

For some organizations, asymmetric distributions pose a problem that VAR on its own *cannot* address, no matter how it is calculated. Consider again the OCIP example, in which the one-year VAR implied a $1 billion loss in one year out of 20. With a symmetric portfolio distribution, that would also imply a $1 billion gain in one year out of 20. But suppose OCIP's investment program had a *positively skewed* return distribution. Then, the $1 billion loss in one year out of 20 might be comparable to, say, a $5 billion gain in one year of 20.

One of the problems with interpreting VAR thus is interpreting the confidence level—*viz.*, 5% or one year in 20. Some organizations may consider it more useful *not* to examine the loss associated with a chosen probability level but rather to examine the risk associated with a *given loss*—the so-called "doomsday" return below which a portfolio must *never* fall. Pension plans, endowments, and some hedge funds, for example, are concerned primarily with the possibility of a "shortfall" of assets below liabilities that would necessitate a contribution from the plan sponsor.

Shortfall risk measures are alternatives to VAR that allow a risk manager to define a specific target value below which the organization's assets must *never* fall and they measure risk accordingly. Two popular measures of shortfall risk are below-target probability ("BTP") and below-target risk ("BTR").[47]

The advantage of BTR, in particular, over VAR is that it penalizes large shortfalls more than small ones.[48] BTR is still subject to the same misinterpretation as VAR when it is reported without a corresponding indication of possible *gains*. VAR, however, relies on a somewhat arbitrary choice of a "probability level" that can be changed to exaggerate or to de-emphasize risk measures. BTR, by contrast, is based on a real target—*e.g.*, a pension actuarial contribution threshold—and thus reveals information about risk that can be much more usefully weighed against expected returns than a VAR measure.[49]

CONCLUSION

By facilitating the consistent measurement of risk across distinct assets and activities, VAR allows firms to monitor, report, and control their risks in a manner that efficiently relates risk control to desired and actual economic exposures. At the same time, reliance on VAR can result in serious problems when improperly used. Would-be users of VAR are thus advised to consider the following three pieces of advice:

First, VAR is useful only to certain firms and only in particular circumstances. Specifically, VAR is a tool for firms engaged in *total value* risk management, where the consolidated values of exposures across a variety of activities are at issue. Dangerous misinterpretations of the risk facing a firm can result when VAR is wrongly applied in situations where total value risk management is *not* the objective, such as at firms concerned with *cash flow* risk rather than value risk.

Second, VAR should be applied very carefully to firms selectively managing their risks. When an organization deliberately takes certain risks as a part of its primary business, VAR can serve at best as a diagnostic monitoring tool for those risks. When VAR is analyzed and reported in such situations with no estimates of corresponding expected profits, the information conveyed by the VAR estimate can be extremely misleading.

Finally, as all the great derivatives disasters illustrate, no form of risk measurement—including VAR—is a substitute for good management. Risk management as a process encompasses much more than just risk measurement. Although judicious risk measurement can prove very useful to certain firms, it is quite pointless without a well-developed organizational infrastructure and IT system capable of supporting the complex and dynamic process of risk taking and risk control.

47. *See* Culp, Tanner, and Mensink, cited previously. For a complete mathematic discussion of these concepts, *see* Kamaryn T. Tanner, "An Asymmetric Distribution Model for Portfolio Optimization," manuscript, Graduate School of Business, The University of Chicago (1997).

48. BTP accomplishes a similar objective but does *not* weight large deviations below the target more heavily than small ones.

49. *See* Culp, Tanner, and Mensink, cited previously, for a more involved treatment of shortfall risk as compared to VAR.

APPENDIX: CALCULATING VAR

To calculate a VAR statistic is easy *once you have generated the probability distribution for future values of the portfolio.* Creating that VAR distribution, on the other hand, can be quite hard, and the methods available range from the banal to the utterly arcane. This appendix reviews a few of those methods.

Variance-Based Approaches

By far the easiest way to create the VAR distribution used in calculating the actual VAR statistic is just to *assume* that distribution is normal (i.e., Gaussian). Mean and variance are "sufficient statistics" to fully characterize a normal distribution. Consequently, knowing the variance of an asset whose return is normally distributed is all that is needed to summarize the risk of that asset.

Using return variances and covariances as inputs, VAR thus can be calculated in a fairly straightforward way.[1] First consider a single asset. If returns on that asset are normally distributed, the 5th percentile VAR is always 1.65 standard deviations below the mean return. So, the variance is a sufficient measure of risk to compute the VAR on that asset—just subtract 1.65 times the standard deviation from the mean. The risk horizon for such a VAR estimate corresponds to the frequency used to compute the input variance.

Now consider two assets. In that case, the VAR of the portfolio of two assets can be computed in a similar manner using the variances of the two assets' returns. These variance-based risk measures then are combined using the correlation of the two assets' returns. The result is a VAR estimate for the portfolio.

The simplicity of the variance-based approach to VAR calculations lies in the assumption of normality. By *assuming* that returns on all financial instruments are normally distributed, the risk manager eliminates the need to come up with a VAR distribution using complicated modeling techniques. All that *really* must be done is to come up with the appropriate variances and correlations.

At the same time, however, by assuming normality, the risk manager has greatly limited the VAR estimate. Normal distributions, after all, are symmetric. Any potential for skewness or fat tails in asset returns thus is totally ignored in the variance-only approach.

In addition to sacrificing the possibility that asset returns may *not* be normally distributed, the variance-only approach to calculating VAR also relies on the critical assumption that asset returns are totally independent across increments of time. A multi-period VAR can be calculated only by calculating a single-period VAR from the available data and then extrapolating the multi-day risk estimate. For example, suppose variances and correlations are available for historical returns measured at the *daily* frequency. To get from a one-day VAR to a T-day VAR—where T is the risk horizon of interest—the variance-only approach requires that the one-day VAR is multiplied by the square root of T.

For return variances and correlations measured at the monthly frequency or lower, this assumption may not be terribly implausible. For daily variances and correlations, however, serial independence is a very strong and usually an unrealistic assumption in most markets. The problem is less severe for short risk horizons, of course. So, using a one-day VAR as the basis for a five-day VAR might be acceptable, whereas a one-day VAR extrapolated into a one-year VAR would be highly problematic in most markets.

Computing Volatility Inputs

Despite its unrealistic assumptions, simple variance-based VAR calculations are probably the dominant application of the VAR measure today. The approach is especially popular for corporate end users of derivatives, principally because the necessary software is cheap and easy to use.

All variance-based VAR measures, however, are not alike. The sources of inputs used to calculate VAR in this manner can differ widely. The next several subsections summarize several popular methods for determining these variances.[2] Note that these methods are *only* methods of computing *variances* on single assets. Correlations still must be determined to convert the VARs of individual assets into portfolio VARs.

1. A useful example of this methodology is presented in Anthony Saunders, "Market Risk," *The Financier*, Vol. 2, No. 5 (December 1995).
2. For more methods, *see* Jorion (1995), cited previously.

For more of a forward-looking measure of volatility, option-implied volatilities sometimes can be used to calculate VAR.

Moving Average Volatility. One of the simplest approaches to calculating variance for use in a variance-based VAR calculation involves estimating a historical moving average of volatility. To get a moving average estimate of variance, the average is taken over a rolling window of historical volatility data. For example, given a 20-day rolling window,[3] the daily variance used for one-day VAR calculations would be the average daily variance over the most recent 20 days. To calculate this, many assume a zero mean daily return and then just average the squared returns for the last 20 trading days. On the next day, a new return becomes available for the volatility calculation. To maintain a 20-day measurement window, the first observation is dropped off and the average is recomputed as the basis of the next day's VAR estimate.

More formally, denote the daily return from time t-1 to time t as r_t. Assuming a zero mean daily return, the moving average volatility over a window of the last D days is calculated as follows:

$$v_t^2 = \left[\frac{1}{D}\right]\sum_{i=0}^{D-1} r_{t-i}^2$$

where v_t is the daily volatility estimate used as the VAR input on day t.

Because moving-average volatility is calculated using equal weights for all observations in the historical time series, the calculations are very simple. The result, however, is a smoothing effect that causes sharp changes in volatility to appear as plateaus over longer periods of time, failing to capture dramatic changes in volatility.

Risk Metrics. To facilitate one-day VAR calculations and extrapolated risk measures for longer risk horizons, JP Morgan—in association with Reuters—began making available their RiskMetrics™ data sets. This data includes historical variances and covariances on a variety of simple assets—sometimes called "primitive securities."[4] Most other assets have cash flows that can be "mapped" into these simpler RiskMetrics assets for VAR calculation purposes.[5]

In the RiskMetrics data set, daily variances and correlations are computed using an "exponentially weighted moving average." Unlike the simple moving-average volatility estimate, an exponentially weighted moving average allows the most recent observations to be more influential in the calculation than observations further in the past. This has the advantage of capturing shocks in the market better than the simple moving average and thus is often regarded as producing a better volatility for variance-based VAR than the equal-weighted moving average alternative.

Conditional Variance Models. Another approach for estimating the variance input to VAR calculations involves the use of "conditional variance" time series methods. The first conditional variance model was developed by Engle in 1982 and is known as the autoregressive conditional heterskedasticity ("ARCH") model.[6] ARCH combines an autoregressive process with a moving average estimation method so that variance still is calculated in the rolling window manner used for moving averages.

Since its introduction, ARCH has evolved into a variety of other conditional variance models, such as Generalized ARCH ("GARCH"), Integrated GARCH ("IGARCH"), and exponential GARCH ("EGARCH"). Numerous applications of these models have led practitioners to believe that these estimation techniques provide better estimates of (time-series) volatility than simpler methods.

For a GARCH(1,1) model, the variance of an asset's return at time t is presumed to have the following structure:

$$v_t^2 = a_0 + a_1 r_{t-1}^2 + a_2 v_{t-1}^2$$

The conditional variance model thus incorporates a *recursive* moving average term. In the special case where $a_0=0$ and $a_1+a_2=1$, the GARCH(1,1) model reduces *exactly* to the exponentially weighted moving average formulation for volatility.[7]

Using volatilities from a GARCH model as inputs in a variance-based VAR calculation does not circumvent the statistical inference problem of presumed normality. By incorporating additional information

3. The length of the window is chosen by the risk manager doing the VAR calculation.

4. Jorion (1995), cited previously.

5. For a detailed explanation of this approach, *see* JP Morgan/Reuters, *RiskMetrics—Technical Document*, 4th ed. (1996).

6. Robert Engle, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol. 50 (1982):391-407.

7. *See* Jorion (1995), cited previously.

into the volatility measure, however, more of the actual time series properties of the underlying asset return can be incorporated into the VAR estimate than if a simple average volatility is used.

Implied Volatility. All of the above methods of computing volatilities for variance-based VAR calculations are based on historical data. For more of a forward-looking measure of volatility, option-implied volatilities sometimes can be used to calculate VAR.

The implied volatility of an option is defined as the *expected future volatility* of the underlying asset over the remaining life of the option. Many studies have concluded that measures of option-implied volatility are, indeed, the *best* predictor of future volatility.[8] Unlike time series measures of volatility that are entirely backward-looking, option implied volatility is "backed-out" of actual option prices—which, in turn, are based on actual transactions and expectations of market participants—and thus is inherently forward-looking.

Any option-implied volatility estimate is dependent on the particular option pricing model used to derive the implied volatility. Given an observed market price for an option *and* a presumed pricing model, the implied volatility can be determined numerically. This variance may then be used in a VAR calculation for the asset underlying the option.

Non-Variance VAR Calculation Methods

Despite the simplicity of most variance-based VAR measurement methods, many practitioners prefer to avoid the restrictive assumptions underlying that approach—*viz.*, symmetric return distributions that are independent and stable over time. To avoid these assumptions, a risk manager must actually generate a full distribution of possible future portfolio values—a distribution that is neither necessarily normal nor symmetric.[9]

Historical simulation is perhaps the easiest alternative to variance-based VAR. This approach generates VAR distributions merely by "re-arranging" historical data—*viz.*, re-sampling time series data on the relevant asset prices or returns. This can be about as easy *computationally* as variance-based VAR, and it does *not* presuppose that everything in the world is normally distributed. Nevertheless, the approach is highly dependent on the availability of potentially massive amounts of historical data. In addition, the VAR resulting from a historical simulation is totally sample dependent.

More advanced approaches to VAR calculation usually involve some type of forward-looking simulation model, such as Monte Carlo. Implementing simulation methods typically is computationally intensive, expensive, and heavily dependent on personnel resources. For that reason, simulation has remained largely limited to active trading firms and institutional investors. Nevertheless, simulation does enable users to depart from the RiskMetrics normality assumptions about underlying asset returns without forcing them to rely on a single historical data sample. Simulation also eliminates the need to assume independence in returns over time—*viz.*, VAR calculations are no longer restricted to one-day estimates that must be extrapolated over the total risk horizon.

8. *See, for example,* Phillipe Jorion, "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance*, Vol. 50 (1995):507-528.

9. Variance-based approaches avoid the problem of generating a new distribution by *assuming* that distribution.

CAS Exam 9 Study Kit

"Allocation of Capital in the Insurance Industry"

by J.D. Cummins

©*Risk Management and Insurance Review*, 2000, Vol. 3, No. 1, 7-27

INVITED ARTICLE

ALLOCATION OF CAPITAL IN THE INSURANCE INDUSTRY

J. David Cummins

ABSTRACT

This article discusses and critiques the methods that have been proposed for allocating capital in financial institutions, with an emphasis on applications in the insurance industry. The author discusses the rationale for allocating capital by line of business and explains how capital allocation can be used to maximize firm value. The implications for capital allocation of regulatory risk-based capital and the capital asset pricing model are discussed. The advantages and disadvantages of using value-at-risk and insolvency put option criteria in capital allocation are analyzed. Finally, recently proposed methods of marginal capital allocation are evaluated. One conclusion is that using the insolvency put option is superior to value-at-risk for allocating capital but that both methods fail to account for diversification across lines in the multi-line firm. The primary conclusion is that marginal capital allocation methodologies based on option-pricing models that recognize the effects of diversification are the best approach for allocating capital in the financial industry.

INTRODUCTION

The purpose of this article is to provide an overview of the various techniques that have been suggested for allocating equity capital. Capital allocation in this context refers to the determination of the amount of a firm's equity capital that is assigned to each project or line of business undertaken by the firm. The objectives of capital allocation are discussed in more detail below. However, firms are usually concerned about capital allocation in the context of pricing and project selection (e.g., determining the proportion of the firm's overall cost of capital that must be contributed by each line of business in order to maximize firm value). The discussion of capital allocation is conducted in the context of the insurance industry. However, most of the techniques discussed are perfectly general and can be applied in other industries as well.

Capital allocation is perhaps of special interest to financial firms such as insurers. For such firms, the principal providers of debt capital (insurance reserves) are also the firm's principal customers. Unlike the holders of bonds and other noninsurance debt capital, insurance policyholders cannot protect themselves against the insolvency of specific

J. David Cummins is Harry J. Loman Professor of Insurance and Risk Management at The Wharton School, University of Pennsylvania, Philadelphia. This article is based on a presentation the author gave at a seminar organized by SCOR Reinsurance and held in Paris in 1998.

debt issuers by holding a diversified portfolio. Unlike the diversified bond investor, the typical policyholder relies on one insurer (or at most a few, in the case of life insurance) for each type of protection purchased (e.g., auto insurance, homeowners insurance, health insurance, etc.). Insurance policies are purchased as protection against adverse financial contingencies. Thus, insolvency risk plays a special role in the insurance industry, and capital is held to assure policyholders that claims will be paid even if larger than expected.

An insurer's entire capital is available to pay the claims arising from any specific policy or line of business. If the insurer becomes insolvent, it is the entire company that enters bankruptcy—the firm does not go bankrupt line by line. Nevertheless, it is often useful to think of capital as being allocated by line of business for pricing, underwriting, and other types of decision making. Looking closely at capital allocation in the insurance industry is also useful to elucidate the interaction between financial decision making and the risk-based capital rules applied to the insurance industry by regulators. Capital allocation is also related to recently emerging concepts such as *risk-adjusted return on capital* (*RAROC*) and *economic value added* (*EVA*), which have become important management decision making techniques for both financial and nonfinancial firms.

Before beginning the discussion of capital allocation, two cautionary notes are in order. First, the terminology in this field is far from standardized. Therefore, a concept such as risk-adjusted return on capital is sometimes called by slightly different names in the literature. It is important to focus on the concepts rather than the nomenclature. Second, because this is an overview, it omits many important details that a firm would have to deal with to allocate capital in practice. Filling in the details to enable insurers to move from the concepts to practical applications in capital allocation provides a promising avenue for future research.

The remainder of the article is organized as follows: the following section discusses how capital allocation can be used to maximize the value of a firm. The next section first provides an overview of the capital-allocation techniques that have been suggested and/or employed in practice and then provides a detailed discussion of the principal methods.

USING CAPITAL ALLOCATION TO MAXIMIZE VALUE

Why allocate capital? The motivation for anything a firm does should be to maximize shareholder value—that is, to increase the market value of equity capital. Unfortunately, this straightforward and powerful objective is often overlooked in practice. Extensive discussions with people in the insurance industry reveal that many firms are managing their GAAP (generally accepted accounting principles) balance sheets and income statements with the objective of showing healthy GAAP earnings and/or maximizing the value of GAAP equity. Of course, the firm needs to be cognizant of its GAAP performance because of the importance of accounting results to financial analysts and traders. However, it is a mistake to lose sight of the firm's true mission—the maximization of market value. Marking to market should play a critical role in the firm's internal decisions.

Capital allocation can be used to facilitate and improve the measurement of the economic profitability of businesses with different sources of risk and different capital requirements. The insurance industry customarily defines businesses in terms of lines of insurance (for example, the commercial liability line or the auto liability line). Although this is the traditional approach, insurers need to think fully about what a business is when designing capital-allocation and performance-measurement systems.

Some authors in the banking literature define deposit accumulation or gaining demand deposits as one business and making loans as another business. In this conceptualization, the economic concept of *transfer pricing* is used, whereby the bank's loan origination business will borrow money from the deposit accumulation business and pay an implicit rate of interest in return for having funds to invest in loans. This approach could also be used in insurance. The underwriting operation could be considered as a funds-generating business in which money is being borrowed from policyholders. The underwriting operation then would lend the funds to the investment business in return for a transfer price.

In either of the preceding lines of business concepts, the maturity and duration characteristics of the debt capital and the investments resulting from the insurer's different businesses must be recognized. Thus, funds generated by issuing long-tail liability policies are likely to lead to different investment objectives than funds raised by issuing short-tail property insurance policies in order to manage the risks of duration and convexity. One cannot assume that the long-term liability line has an asset portfolio that looks like the company as a whole; it has to be managed to meet the firm's overall objectives in terms of interest rate risk. Duration and convexity management is extremely technical. Therefore, special care must be exercised in giving a particular business credit for the money it generates while, at the same time, charging it for the use of capital. This is the case for both the debt and equity capital needed to operate the business.

The primary link between capital allocation and value maximization is to enable the firm to measure performance by line of business to determine whether each business is contributing sufficiently to profits to cover its cost of capital and add value to the firm. To measure the cost of capital, it is necessary to determine the capital required to offer each type of insurance—relatively risky lines typically require more capital than less risky lines. For example, one might ask whether the commercial liability insurance business is making an adequate profit—that is, whether the insurer should be charging higher or lower prices than at present and whether it should exit this business or perhaps devote more capital to it. Insurers can maximize value by shedding unprofitable businesses as well as by identifying profitable new projects. By withdrawing from unprofitable lines, the insurer may be able to increase the market value of equity, even as revenues decline. The ultimate objective should not be revenue growth, but maximizing net worth.

To provide a framework for the discussion of capital allocation methodologies, it is helpful to provide a simple mathematical statement of the capital allocation problem. Define x_i as the proportion of the firm's equity capital allocated to business i, where x_i is between 0 and 1. Thus, x_i indicates the proportion of capital that is allocated to business i and, therefore, the amount of capital allocated to business i is C_i, which is the total capital, C, multiplied by x_i. If the firm has N businesses, then

$$\sum_{i=1}^{N} x_i \leq 1 \text{ and } \sum_{i=1}^{N} C_i \leq C.$$

That is, the sum of the capital allocated to all of the firm's business will be less than or equal to the firm's total capital. Although it may seem surprising that a firm may not assign all of its capital to its businesses, some leading-edge researchers argue against allocating all of the capital and favor, instead, an allocation that results in less than 100 percent being assigned (Merton and Perold, 1993). This issue is addressed in the discussion below of the use of option models to allocate capital.

Once capital has been allocated by line, how can the resulting allocations be used to maximize firm value? One approach that has received considerable attention is to calculate the *risk-adjusted return on capital (RAROC)*. *RAROC* is defined as the net income from a line divided by the capital allocated to the line. That is,

$$RAROC_i = \frac{Net\ Income_i}{C_i},$$

where C_i is the capital allocated to line of business *i*. The numerator of the *RAROC* formula, net income, also needs to be defined carefully. It may seem obvious but is not once it is considered in economic terms. Basically, net income in the *RAROC* formula should be after taxes and interest expense. Even though interest expense is a banking term, it also applies to insurers in the form of underwriting loss. That is, the insurance market implicitly discounts the loss cash flows for the time value of money, meaning that the underwriting profit is negative in most cases. The negative underwriting return, which is analogous to interest expense, needs to be taken out when calculating the return from a line of business.

Once the *RAROC* for a line of business has been calculated, how does the firm know whether the line's current risk-adjusted return is adequate? The risk-adjusted return should be compared with the cost of capital for business *i*, where the cost of capital is obtained using an appropriate asset pricing model. If the risk-adjusted return equals or exceeds the cost of capital, then continuing to devote resources to this line of business is consistent with the goal of value maximization. However, if the risk-adjusted return is below the cost of capital, the line of business is reducing the firm's market value. In this circumstance, the firm should take some action to improve the situation, such as repricing the insurance, tightening underwriting standards, or withdrawing from the line of business.

A slightly more formal way of determining whether a particular line of business is adding to firm value goes under the name of *economic value-added (EVA)*.[1] Economic value-added measures the return on an investment in excess of its expected or required return. *EVA* seeks to identify lines that create value for the firm. *EVA* is net income minus the cost of capital, or hurdle rate, for a certain business, multiplied by the capital allocated to the business:

$$EVA_i = Net\ Income_i - r_i C_i,$$

where r_i = cost of capital (hurdle rate) for business *i*. Thus, if $EVA \geq 0$, writing the line of business is consistent with value maximization, while if $EVA < 0$, the line is eroding firm value. The *EVA* formula can be changed slightly to put the results in rate of return format, creating a measure called the *economic value added on capital (EVAOC)*. *EVAOC* is defined as *EVA* divided by the capital allocated to a line:

$$EVAOC_i = (Net\ Income_i / C_i) - r_i.$$

This is similar to *RAROC*, except that the line's cost of capital is subtracted. Again, if *EVAOC* is positive, the line is creating value for the firm.

[1] Several other concepts similar to *EVA* have been discussed in the literature. The author uses *EVA* as an example of these concepts.

An important detail is how to determine the cost of capital for business *i*. This too can present a problem in the insurance industry because of data limitations. One approach proposed by finance researchers to estimating the cost of capital for a line of business is the "pure play" technique. The pure play approach estimates the cost of capital by finding other firms that offer only one line of business. The cost of capital for a business in the multiple-line firm can then be based on the cost of capital of mono-line firms offering only that specific type of business. This approach is problematic in the insurance industry because it is difficult to find firms that write only one line of business. Even if such firms could be found, the underwriting risk characteristics of the pure play firms could differ significantly from those of a given line of business written by a multiple-line firm. An alternative to the pure play technique is the use of "full-information betas" to determine the cost of capital (Kaplan and Peterson, 1998). This estimation technique uses data on conglomerates (firms that offer several lines of business) to conduct regressions that permit the estimation of the cost of capital by line.

Another problem in the insurance industry is the lack of quality data. An insurer thinking about implementing *value at risk* (*VaR*), *EVA*, and the other economic methodologies discussed here needs to think about revising its data system to capture the data required to implement the methodologies. Insurers should be designing data systems that allow them to report underwriting results frequently (for example, on a monthly basis). Data quality is crucial. With inadequate data, even a perfect model will fail. Currently, most insurers do not have the necessary data to implement *VaR*, *RAROC*, and *EVA* effectively.

CAPITAL ALLOCATION TECHNIQUES

Various methodologies have been developed that could provide the foundation for a system of capital allocation. The following is by no means an exhaustive list. Many other proposals can be found throughout the related literature. The article first provides an overview of the methods and then discusses each method separately.

Overview of Capital Allocation Techniques

Regulatory Risk-Based Capital In the United States, regulators have developed a formula to calculate the risk-based capital (RBC) of insurers. A company's risk-based capital is used to define the minimum capital it must hold in order to avoid regulatory intervention. The regulatory thresholds ("action levels") are determined by the risk-based capital ratio, which is the ratio of the company's actual capital to its risk-based capital (see Cummins, Harrington, and Niehaus, 1993). If the insurer's actual capital is greater than 200 percent of its risk-based capital, no regulatory action occurs. However, if actual capital falls below 200 percent of risk-based capital, various regulatory actions are taken, depending upon how far actual capital falls below 200 percent. An insurer's risk-based capital is computed by a formula designed to require more capital for riskier companies.

Some insurers use the regulatory risk-based capital formula in allocating capital for purposes of managing the firm. In the author's opinion, this is unwise, because the regulatory charges are of questionable accuracy and are based on book rather than market values.[2] Furthermore, the regulatory charges ignore important sources of risk such as

[2] Cummins, Harrington, and Klein (1995) show that risk-based capital has low predictive power in terms of identifying companies that are likely to become insolvent.

interest rate (duration and convexity) risk, as well as the insurer's transactions in the derivatives market. Even if the charges were accurate, they would be accurate only for the average firm in the industry. Consequently, in the case of firms with books of business having above- or below-average risk, the regulatory charges would produce inappropriate allocations of capital. The result is that businesses may be charged for too much or too little capital, leading to suboptimal decisions.

The Capital Asset Pricing Model The second approach to capital allocation involves using one of the oldest of modern financial theory technologies: the capital asset pricing model (CAPM). This is not the best solution to the problem, but it may provide a helpful benchmark based on a familiar methodology. At the very least, the use of the CAPM allows managers to compare the preferred method to the results generated by a classic technique. A similar analysis could be conducted using more advanced asset pricing models.

Value at Risk The value at risk concept has become very popular in the banking and investment banking communities, where there is a need to examine the risk exposure of the firm's trading book for foreign exchange, bonds, etc. Simply stated, *VaR* is the amount the firm could lose with a specified small probability, such as 1 percent, in a specified period of time. The measurement of value at risk from currency and securities trading has advanced rapidly, in part because of daily and even more frequent data on exchange rates and asset prices that allow for very accurate and sophisticated calculations of *VaR*.[3]

VaR is also likely to be useful to insurers and, in fact, is closely related to time-honored insurance and actuarial concepts such as the probability of ruin and the maximum probable loss. Unfortunately, the application of the most sophisticated *VaR* techniques requires very frequent data (monthly data is an absolute minimum), but insurance prices and losses are not observed with sufficient frequency nor in a market context (although most insurers do generate such data internally). Using sophisticated tools such as value at risk requires an integration of the capital allocation methodology with data processing and information systems to ensure that pertinent and useful data are generated to provide inputs for the *VaR* models.

Marginal Capital Allocation Marginal capital allocation is a term that can be applied to the capital allocation technique proposed by Robert Merton and Andre Perold (1993) and to a related technique developed by Stewart Myers and James Read (1999). Both techniques are based on the option pricing model of the firm. In the options view of the firm, the value of the policyholders' claim on the firm is equal to the present value of losses minus the value of the *insolvency put option*. The insolvency put option is the expected loss to policyholders because of the possibility that the firm will default. The simplest option interpretation of the firm involves a one-period, two-date model in which the firm issues policies at time 0 and claim payments occur at time 1. If assets exceed liabilities at time 1, the firm pays the losses and the equity owners receive the residual value (the difference between assets and liabilities). However, if assets are less

[3] Some sophisticated academic research also has been developed on *VaR*. One important problem is to allow for time-varying volatility. For example, the volatility of foreign exchange markets may be very high in one period and very low in another period. Another important issue is to allow for autoregressive patterns in the data, whereby volatility in one period may depend on things that have occurred in prior periods.

than liabilities, the insurer defaults and the policyholders receive the assets. The payoff to policyholders at time 1 is thus equal to : $L - \text{Max}[L - A, 0]$, where L = losses, A = assets, and $\text{Max}[L - A, 0]$ is the payoff on the insolvency put.

The Merton-Perold (M-P) and Myers-Read (M-R) capital allocation techniques take different views of what is meant by a "marginal" approach to capital allocation. Stated simply, the M-P approach views marginal capital allocation in terms of what happens to the insolvency put option if entire lines of business are added to or removed from the firm. The M-R approach is marginal in the instantaneous interpretation familiar from calculus; i.e., they allocate capital based on what happens to the firm's insolvency put option in response to very small changes in the liabilities of the lines of business written by the firm. The M-P method may leave some capital unallocated, while the M-R method makes a unique allocation of 100 percent of capital.

Regulatory Risk-Based Capital (RBC)

The author's examination of regulatory RBC is based on the RBC system used in the United States. RBC has been in effect for life insurers since 1993 and for property-liability insurers since 1994. Even though it is not appropriate to use regulatory capital as a management tool, the regulatory approach is important because it identifies most of the important risks faced by insurers. In addition, regulatory capital may serve as a constraint on certain of the firm's activities, because violation of regulatory solvency boundaries can subject the firm to regulatory costs.

The risk-based capital systems consist of RBC proportions, which are multiplied by various income statement and balance sheet quantities to generate "RBC charges." The sum of the charges equals an insurer's risk-based capital, following a "covariance adjustment" discussed below. The following are the principal risk-based capital charges for property-liability insurers:

- Investment RBC = R_1

  Various charges exist for bonds, stocks, and other investments. For example, the charge ratio is zero for treasury bonds, then progressively higher, capping at 0.3 for bonds in or near default.

- Loss Reserve RBC = R_2

  Loss reserve RBC allows for the risk of adverse reserve development (i.e., the possibility that loss payments will exceed reserved amounts). The proportions vary by line that are multiplied by outstanding loss reserves.

- Written premium RBC = R_3

  Written premium RBC accounts for the possibility that the loss ratio on the new business will be higher than expected. The charges vary by line and are multiplied by net premiums written.

- Credit RBC = R_4

  Credit RBC allows for the possibility that agents and reinsurers will default on their obligations to pay the company the money they owe.

- Off-balance sheet RBC = R_5

Off-balance sheet RBC allows for unexpected payments from contracts that do not show up on the balance sheet such as loan guarantees for subsidiaries. There is no charge for derivatives at the present time, even though derivative transactions may pose a significant off-balance sheet risk for some insurers.

After the risk-based capital charges are determined, they could be added to produce the company's total risk-based capital. However, this would overlook diversification (i.e., it would not ignore the likelihood that an insurer will not be affected by adverse outcomes for all of the risks simultaneously). Indeed, it is possible that the risks facing insurers are less than perfectly correlated; that is, an adverse outcome from one type of risk could be offset by a favorable outcome from another. Because determining the correlations among risks is difficult, the RBC system assumes that the correlations are zero; i.e., that the risks are uncorrelated. This assumption provides the foundation for the RBC formula:

$$R_T = R_0 + \left[R_1^2 + R_2^2 + R_3^2 + R_4^2 + R_5^2\right]^{1/2},$$

where R_0 = risk-based capital for holdings of stocks of the firm's subsidiaries and R_1 through R_5 are defined earlier. The square root in the formula is a *covariance adjustment* implicitly assuming zero correlation among risks 1 through 5.

Why would it be a mistake for insurers to use RBC in capital allocation? One reason is that there is no theoretical foundation for the formula. Using a purely empirical approach with no theory would be a mistake. Second, the formula is of questionable accuracy (Cummins, Harrington, and Klein, 1995). One of the defects of the model is that some of the charges are based on worst-case scenarios; e.g., adverse historical outcomes, rather than using statistical concepts such as variances and covariances to estimate relative risks. The formula also ignores correlations among the firm's businesses, which should be taken into account in an accurate capital-allocation system.

The Capital Asset Pricing Model (CAPM)

The CAPM states that the expected return on equity or cost of capital for a firm is determined by the following formula:

$$r_E = r_F + \beta_E(r_M - r_F),$$

where r_E = cost of equity capital,

r_F = default risk-free rate of interest,

r_M = expected return on the "market," and

β_E = the firm's beta coefficient = $\mathrm{Cov}(r_E, r_M)/\mathrm{Var}(r_M)$.

where Cov(•) = the covariance operator and Var(•) = the variance operator.

How can the CAPM be used by an insurance company to make pricing and investment decisions? Project decisions can be made by decomposing the beta coefficient to determine the betas by line of business. For example, let's consider an insurer with two lines of business. Its net income would be:

$$I = r_A A + r_1 P_1 + r_2 P_2,$$

where I = net income,
r_A = return on assets,
r_1, r_2 = rates of return on underwriting from lines 1 and 2,
A = assets, and
P_1, P_2 = premiums from lines 1 and 2.

Next introduce the balance sheet identity, which says that assets are equal to equity, plus the liabilities generated by the two lines. Then divide by equity to express the result as a rate of return:

$$r_E = r_A(E + L_1 + L_2) / E + r_1 P_1 / E + r_2 P_2 / E.$$

Then the beta coefficient can be decomposed as follows:

$$\beta_E = \beta_A(1 + k_1 + k_2) + \beta_1 s_1 + \beta_2 s_2,$$

where $\beta_E, \beta_A, \beta_1, \beta_2$ = betas for the firm, assets, and insurance risk of lines 1 and 2,
k_1, k_2 = liability leverage ratios for lines 1 and 2, $= L_i/E$, $i = 1, 2$, and
s_1, s_2 = premium leverage ratios for lines 1 and 2, $= P_i/E$, $i = 1, 2$.

The calculation uses the property that the covariance operator is linear.

The formula for the decomposition of β_E shows that the beta of the firm, which drives the required return on equity, is the beta of assets times 1, representing the investment of equity capital, plus the liability leverage ratios for lines 1 and 2. Then the formula adds on the beta of each individual line's underwriting returns multiplied by the line-specific premium-to-surplus ratio. This is a theoretical justification for the traditional rule-of-thumb leverage ratio that has been used for years in the insurance industry—the premium-to-surplus ratio.

The model can be solved for the required rate of underwriting return on each line of business:

$$r_i = -k_i r_F + \beta_i(r_M - r_F)$$

for lines of business $i = 1$ and 2. Thus, each line implicitly pays interest for the use of policyholder funds (the term $-k_i r_F$) and receives a rate of return based on the systematic risk of the line [the term $\beta_i (r_M - r_F)$].

The CAPM result has the following implication: It is not necessary to allocate capital by line using the CAPM, but rather to charge each line for at least the CAPM cost of capital, reflecting the line's beta coefficient and leverage ratio. Costs of capital based on other asset pricing models, such as the arbitrage pricing theory (APT), have similar implications.

Although the CAPM provides a useful way of conceptualizing the contributions of the firm's lines of business to the return on equity, there are at least three important problems with this model.

1. The CAPM only rewards the firm for bearing systematic underwriting risk; that is, underwriting risk that is correlated with the market portfolio. However, insurers also need to be concerned about extreme events—for example, tail risk—that are simply not priced in the CAPM model. This is important in view of the role of insurers as financial intermediaries, where a firm's principal creditors are also its customers.

2. Line of business underwriting betas are difficult to estimate given the data currently available (Cummins and Harrington, 1988), although progress has been made in estimating costs of capital in the more recent literature (e.g., Lee and Cummins, 1998).

3. Research has shown that rates of return are driven by other economic factors besides beta coefficients (e.g., Fama and French, 1996). Thus, sole reliance on the CAPM would ignore important determinants of the cost of capital. The primary role for the CAPM is to serve as a benchmark to compare with the results of other estimation methodologies. If the two methods yield drastically different results, it would be appropriate to check the methodology and data carefully before proceeding.

Value at Risk (*VaR*)

VaR is defined as the maximum amount the firm could lose over a specified time period with a specified (usually very small) probability. For example, a currency trader might want to know how much could be lost in one week with a probability of 1 percent. For an insurance company, it might be sufficient to ask how much could be lost in a calendar quarter. For example, at the 1 percent probability level, the maximum amount a given insurer can lose on a given line of business over the next six months might be estimated as $5.6 million. The idea is to make certain the insurer's managers understand the degree to which a given line of business is exposing the firm to potentially large losses at any given time.

VaR has become very important in industries such as banking. Although some bank researchers think they have invented something new, in fact, concepts similar to value at risk have been used for a long time in the insurance industry. For example, actuaries have been estimating ruin probabilities for many years; and *VaR* is also similar to the maximum probable loss concept, which is widely used by reinsurers, insurance brokers, and risk managers. The methodologies needed to compute business-specific and firm-wide *VaR* already exist in both the actuarial literature and the financial literature.

How can *VaR* be used in capital allocation? One approach is to use *exceedence probabilities.* The exceedence probability is defined as the probability that losses from a particular line of business will exceed the expected loss plus the capital allocated to the line. Representing the exceedence probability by ε, it can be defined more formally, as follows:

The Exceedence Probability (ε):

$$\text{Probability}\,[Loss_i > \mathrm{E}(Loss_i) + C_i] = \varepsilon,$$

where $Loss_i$ = loss from insurance line i,

$\mathrm{E}(Loss_i)$ = expected value of loss from line i, and

C_i = capital allocated to line i

Capital can be allocated by equalizing the exceedence probabilities across the lines of business written by the insurer. For example, for two lines, we would have:

$$\text{Probability } [Loss_1 > \mathrm{E}(Loss_1) + C_1] = \varepsilon = \text{Probability } [Loss_2 > \mathrm{E}(Loss_2) + C_2].$$

If the lines of business have expected losses that differ in size, it is convenient to express the result in terms of ratios to expected losses:

$$\Pr\,[Loss_1/\mathrm{E}(Loss_1) > 1 + C_1/\mathrm{E}(Loss_1)] = \varepsilon = \Pr\,[Loss_2/\mathrm{E}(Loss_2) > 1 + C_2/\mathrm{E}(Loss_2)],$$

where the abbreviation Pr [•] is used to represent probability. This formula expresses the required capital in terms of ratios of capital allocations to the expected loss of each line of business. Lines of business with relatively high risk would require more capital relative to expected losses in order to attain the specified exceedence probability. Exceedence probability curves are shown in Figure 1, which plots the amount of loss on the horizontal axis against the probability of loss on the vertical axis. The ratios of expected losses plus capital needed to achieve a 5 percent exceedence probability in each of the three lines can be determined from the graph. The ratio for the most risky line (line 3) is higher than the ratios for the less risky lines (lines 1 and 2). The ratio for line 3 (2.8) implies that the insurer would have to commit $1.8 in capital for each dollar of expected losses in this line to achieve the target 5 percent exceedence probability.

Three issues to think about in terms of capital allocation using *VaR* are the following:

1. The firm may not have enough capital to attain a given exceedence probability for all of its businesses. In this case, it can either raise the exceedence probability or raise more capital.
2. The exceedence probability approach as outlined here does not consider the diversification effect across lines. This diversification effect can be taken into account using another method discussed later.
3. The exceedence probability does not tell managers anything about the *amount* by which losses are likely to exceed the available resources if an event occurs that breaches the exceedence level.

The latter two issues are addressed by the methodologies discussed next.

The Insolvency Put Option

Option pricing theory can be used to allocate capital in a manner that is conceptually similar to but more general than value-at-risk analysis. The allocation is conducted using the option model of the firm. Consider an insurance firm with random assets (A) and liabilities (L) that is subject to default risk. The liabilities (policyholder claims) are assumed to be payable one period in the future (this is a European rather than an American option example). If assets exceed liabilities ($A > L$) at the liability maturity date, the policyholders receive the value of the liabilities and the owners or residual claimants receive the remaining assets of the firm. However, if the value of the assets is less than the value of the liabilities ($A < L$), the owners of the firm default on the liability payment and the policyholders receive the assets of the firm.

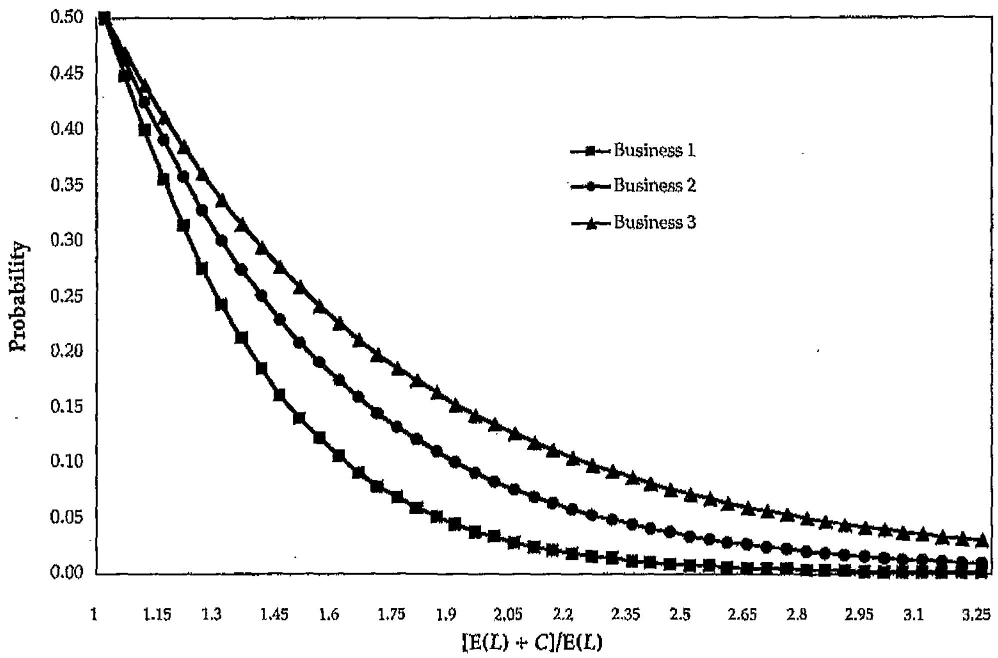
To allocate capital, consider the value of the policyholders' claim on the firm before the maturity of the liabilities (that is, economic value of the policyholders' claim on the firm before the liability maturity date). The value of this claim is the risk-less present value of the liabilities, which would be the present value of the liabilities if default risk were zero, minus a "put option," which is a function of assets (A), liabilities (L), the risk-free rate (r), the time to maturity (τ), and volatility (σ). That is, the value of the policyholders' payoff is

$$\text{Value of Policyholders' Claim} = L\,e^{-r\tau} - P(A,L,r,\tau,\sigma),$$

where $P(A,L,r,\tau,\sigma)$ = value of a put option on A with strike price L, interest rate r, time to maturity τ, and risk parameter σ. The risk parameter σ reflects the volatility of assets and liabilities, as well as the correlation between them. $P(A,L,r,\tau,\sigma)$ is called the *insolvency put option* or the *expected policyholder deficit (EPD)* (Butsic, 1994).

Capital can be allocated based on the insolvency put option. This approach is similar to, but better than, the value-at-risk methodology because it considers the expected value of the amount that can be lost rather than just giving the probability of exceeding a specified value of loss. In order to illustrate this approach to capital allocation, consider three lines of business with different levels of risk, where the level of risk is determined by the value of the option risk parameter, σ. Here, there are three lines of business: a low-risk line with σ of 0.375, an intermediate-risk line with σ of 0.5, and a high-risk line with σ of 0.625.[4]

FIGURE 1
VaR Exceedence Probability Curves

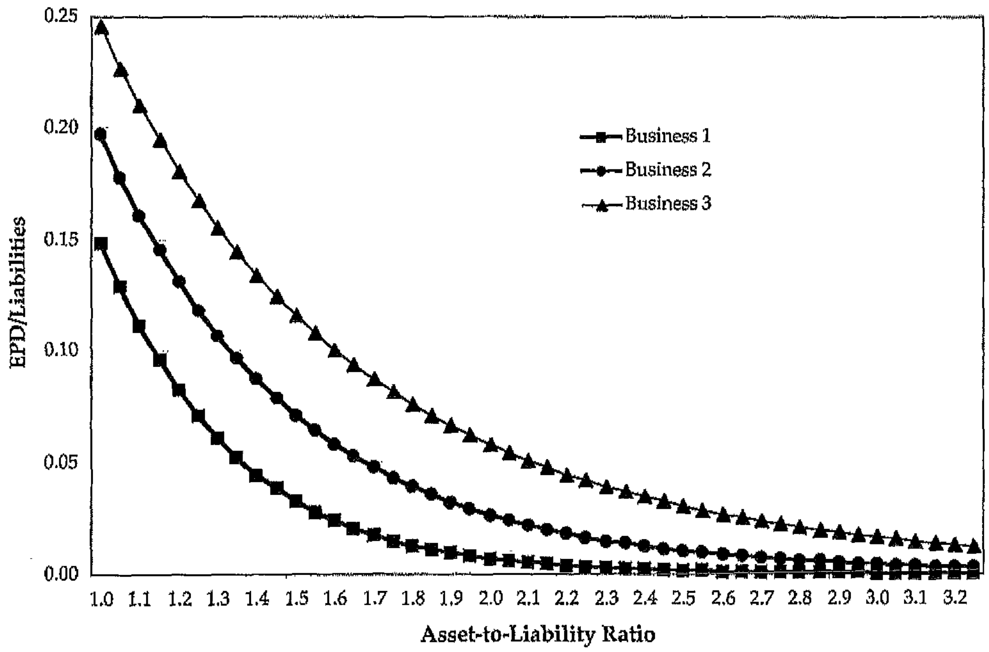


[4] In terms of the put option pricing function $P(A,L,r,\tau,\sigma)$, the other parameters used to compute the option values shown in Figure 1 are: $r = 0$, $\tau = 1$, $L = 1$, and A = asset-to-liability ratio. The example is based on one presented in Merton and Perold (1993). The Black-Scholes option pricing model is used in the example. The numbers shown on the graphs have been rounded.

This approach allocates capital to each line such that the insolvency puts or expected policyholder deficits, expressed as a proportion of liabilities, L, are equalized at a specified target level, such as 5 percent. To illustrate the method, the put option values for the three businesses are shown in Figure 2 as a function of the asset-to-liability ratio. The asset-to-liability ratio measures the capital assigned to each line because assets (A_i) are equal to L_i (the nominal loss liability for line i) + C_i (the capital allocated to line i), so the asset-to-liability ratio equals $[1+C_i/L_i]$. As shown in the chart, the value of the EPD, or put option, for each line declines as the amount of capital allocated to the line increases.

To continue with the capital allocation example, assume that the firm has selected a target EPD of 5 percent of liabilities. To achieve this target, an asset-to-liability ratio of 1.4 is required for business 1 (the low-risk line of business). Or, in other words, $0.4 of capital must be allocated to the line for each $1 of nominal loss liability. Similarly, the intermediate-risk line requires an asset-to-liability ratio of 1.7, and the high-risk line requires an asset-to-liability ratio of 2.1. Thus, the amount of capital needed to achieve the EPD objective is directly related to the risk of the line, as measured by the volatility parameter, σ.

FIGURE 2
Expected Policyholder Deficit



The EPD, or "insolvency put," allocation method is similar to the *VaR* method. However, the EPD is more informative than *VaR* because it considers the expected amount of loss that will occur with a specified probability rather than just the amount of loss that will be exceeded with a specified probability. For example, it is more useful to know that the expected loss in excess of some asset-to-liability ratio (e.g., 2.1) will be $5 million rather than just to know that the probability of exceeding the asset-to-liability ratio is 0.05. The firm would presumably react differently if the expected amount by which the asset to liability ratio is exceeded is $100 versus $100 million. (Of course, for most purposes, it would be useful to know both the *VaR* and the EPD.) The EPD also has the advantage of being consistent with the financial theory of pricing risky debt contracts. This is important because financial theory tells us how to maximize value, or add value to the firm.

The EPD capital-allocation method represents an improvement over value at risk. However, the method should not be taken to imply that the lines of business are separate from each other in terms of default risk. All lines of business still have access to the firm's entire equity capital, and the firm goes bankrupt overall and not by line. Nevertheless, the method does provide some valuable information as long as the insolvency issue is kept in mind. However, the EPD approach has a serious defect; i.e., it does not take into account the benefits of diversification across business lines.

Recognizing Diversification: Marginal Capital Allocation

It is important to recognize the effects of diversification when allocating capital and measuring returns. An insurer can reduce its overall risk by engaging in a portfolio of businesses whose returns are not perfectly correlated.

Consider first the case in which the businesses are all separated into distinct companies, as they would be if a multiple-line firm were to spin off its businesses into separate entities. The spinoff businesses would need more total capital to achieve a given *VaR* or EPD goal than would the multiple-line firm. The multiple-line firm would need less capital because of diversification across lines. A particular line may have high stand-alone risk, but its return may have a low correlation with returns on other lines of business, making the line very valuable from a diversification perspective. Treating the line on a stand-alone basis would lead to excessive allocations of capital to the line and the underestimation of its risk-adjusted return on capital. Many of the existing capital allocation techniques do not take diversification into account.

How can diversification benefits be recognized when allocating capital? An option-based capital allocation method that recognizes the benefits of diversification has been proposed by Merton and Perold (M-P) (1993). They base their analysis on the concept of *risk capital*, which is defined as follows:

> Risk Capital is the smallest amount that can be invested to insure the value of the firm's net assets against a loss in value relative to the risk-free investment of those net assets.

The firm's net assets are defined as its gross or total assets minus the default-risk-free value of the firm's policyholder liabilities. Risk capital can be thought of as the value of an option purchased by the firm that guarantees that the firm will receive no less than its net assets at the option maturity date. Therefore, if the firm has gross assets (A) of $1,000 and the default-free value of its liabilities is $500, the option pays off if the gross value of assets falls below $1,000. Risk capital can come from various sources. It could be totally supplied by the firm. However, if the firm has default risk, risk capital is partially supplied by the liability holders.

The Merton-Perold methodology can be described by continuing the example used above to illustrate the allocation of capital by expected policyholder deficits. Again consider the three businesses whose put option values are plotted in Figure 2. All businesses are assumed to have nominal liabilities of $1,000, and the volatility parameters used in the illustration are the same as before. However, to consider diversification, assumptions are also needed about the correlations among the lines of business. In the context of the Black-Scholes option model used in the illustration, the relevant correlations are between the natural logarithms of the loss liabilities. Assume the following log correlations: $\rho_{12} = 0.5$, $\rho_{13} = 0.75$, and $\rho_{23} = 0.5$, where ρ_{ij} = log correlation between line i and line j.

To illustrate marginal capital allocation, first consider the stand-alone capital required for the three businesses. Stand-alone capital is the capital that would be needed by the firm's businesses if they were operated as separate firms. Assume that all three businesses adopt an expected policyholder deficit target of 5 percent of the nominal (risk-free) value of liabilities. The stand-alone capital for the three firms is shown in Table 1.

TABLE 1
Stand-Alone Capital

EPD = 5 percent of liabilities			
Business	*Assets*	*Nominal Liabilities*	*Stand-Alone Capital*
1	1,361	1,000	361
2	1,672	1,000	672
3	2,107	1,000	1,107

$\sigma_1 = 0.375;\ \sigma_2 = 0.5;\ \sigma_3 = 0.625;\ \rho_{12} = 0.5;\ \rho_{13} = 0.75;\ \rho_{23} = 0.5;\ r = 0;\ \tau = 1.$

As Table 1 shows, business 1 requires assets of $1,361 to attain the 5 percent EPD objective, while businesses 2 and 3 require assets of $1,672 and $2,107, respectively. These are the same values portrayed at the 5 percent level for the three businesses in Figure 2, although the numbers in the figure have been rounded for presentation purposes.

One could combine the three businesses into a single firm and add the stand-alone risk capital of each of the three businesses to obtain the risk capital of the entire enterprise. However, because of diversification, the total amount of capital needed in order for the single enterprise to attain the target EPD is less than the sum of the stand-alone capital for the individual lines. In fact, the capital required to achieve the 5 percent EPD target if the three businesses were combined into a single firm would be $1,427, significantly less than the sum of the stand-alone capital of the three businesses ($2,140). The question is how to distribute the benefits of diversification among the three businesses.

The M-P method of capital allocation is conducted in two steps:

1. Calculate the risk capital required by firms that combine two of the businesses. Here, there are three possible combinations: businesses 1 and 2, businesses 1 and 3, and businesses 2 and 3.
2. Calculate the *marginal* capital required when the excluded business is added to the two-business firms (i.e., this is the marginal capital required if a firm operating two businesses were to add the third business).

The capital allocated to a given business is equal to the marginal capital required when it is added to the appropriate two-business firm. Because the calculation is made for each two-firm combination, the method provides a capital allocation for each of the three businesses comprising the firm. The order in which the businesses are combined into firms does not matter because all three two-business combinations are used; i.e., the allocated capital of each business is obtained on the assumption that two of the businesses have already been combined.

To illustrate M-P capital allocation, first consider the firms obtained by forming two-business combinations. The capital required to achieve the 5 percent EPD target for the two-business firms is shown in Table 2. The table shows the amount of capital needed to achieve the EPD target if businesses 1 and 2 were combined to form a single firm, the amount required if businesses 1 and 3 were combined, and the amount required if businesses 2 and 3 were combined. For each two-business combination, the amount of capital needed for the combined firm is less than the sum of the stand-alone capital of the two businesses comprising the combined firm. This effect occurs because the lines of business are not perfectly correlated.

TABLE 2
Joint Versus Stand-Alone Capital
Example: Effects of Diversification

Business Lines	*Joint Capital*	*Total Stand-Alone Capital*
1 & 2	745	1,033
1 & 3	1,175	1,468
2 & 3	1,276	1,779

The second step in the M-P capital allocation method is to calculate the additional capital required if the excluded businesses are added to the two-business combination firms shown in Table 2. This step is illustrated in Table 3, which shows the additional capital needed when each business is considered to be the marginal case. The marginal allocations, shown in the column headed "Merton-Perold Marginal Capital," are obtained by subtracting the capital needed for the two-line firm, not including business line *i*, from the amount of capital needed by the firm consisting of all three lines of business ($1,427). For example, consider the two-business firm consisting of businesses 1 and 2. This firm consisting of lines 1 and 2 needs only $745 in capital to achieve a 5 percent EPD (see Table 2). Adding business 3 to this firm nearly doubles the required amount of capital. The marginal capital of business 3 is equal to the total capital required for the three-line business to achieve the 5 percent EPD target ($1,427) minus the amount of capital needed by the two-line firm consisting of businesses 1 and 2 ($745), or $682 = $1,427 – $745. The marginal capital of firms 1 and 2 is computed similarly. The capital allocations to the three lines are directly related to the risks of the lines, with line 3 being given the largest capital allocation because it has the highest risk.

Table 3 also shows that the Merton-Perold marginal capital allocation assigns less than 100 percent of the three-line firm's capital to the three businesses. The total amount allocated is $1,084, which is $343 less than the three-line firm's total required capital of $1,427. It is important to emphasize that the three-line firm requires the entire amount of capital ($1,427) in order to attain the EPD target of 5 percent. However, the amount of capital allocated to the three businesses marginally is less than the total capital of $1,427. This means that some of the firm's capital is allocated to the "corporate" level rather than to any of the three individual businesses.

TABLE 3
Allocated and Unallocated Capital

Current Lines	*Business Added*	*Merton-Perold Marginal Capital*	*Stand-Alone Capital*	*Myers-Read Marginal Capital*
1 & 2	3	682	1,107	811
1 & 3	2	252	672	392
2 & 3	1	150	361	224
Totals		1,084	2,140	1,427
Risk Cap for 1, 2, & 3		1,427		1,427
Total Allocated		1,084		1,427
Unallocated		343		0

Merton and Perold (1993) argue that capital allocations based on stand-alone capital are likely to lead to incorrect decisions about the projects undertaken by the firm and the performance evaluation of lines of business. As the above example suggests, the M-P method implies that marginal risk capital should be used in the denominator in calculating performance statistics such as the risk-adjusted return on capital (*RAROC*) and economic value added (*EVA*) for a firm's businesses. Unless the businesses are perfectly correlated (in which case they are actually the same business), this approach results in higher estimates of *RAROC* and *EVA* than if the total amount of the firm's capital were allocated to the businesses. M-P argue that a full allocation of capital will lead the firm to reject projects that would add to its market value. Their view is that adoption of their marginal approach will lead the firm to make value-maximizing decisions, while use of another approach could lead the firm to avoid some projects that would add to the firm's market value of equity.

Myers and Read (M-R) (1999) also use an option pricing model to allocate capital. However, they reach different conclusions from Merton and Perold. Whereas Merton and Perold allocate capital at the margin by adding entire businesses to the firm (a *macro* marginal allocation), Myers and Read allocate capital by determining the effect of very small changes in loss liabilities for each line of business (a *micro* marginal allocation). Myers and Read show that their method leads to a unique allocation of capital by line that assigns 100 percent of total capital.

The M-R approach can be illustrated by continuing the above example using the hypothetical firm investigated in this article.[5] Define the firm's default value (insolvency put option) as $P(L_1, L_2, L_3)$, where $P(\bullet)$ = the insolvency put and L_i = loss liabilities for line i, i = 1, 2, 3, valued at date zero. They then derive capital allocation formulas by taking the

[5] Myers and Read (1999) assume that the firm invests in risky assets and therefore allow for asset volatility as well as covariability between assets and the losses of the three businesses. To keep the example as simple as possible, while focusing on the essential concepts, the author assumes that the firm invests in risk-free assets. The conclusions would be the same if asset risk were present in the example.

derivative of P with respect to each of the L_i. Thus, instead of adding entire businesses to the firm, they perform an incremental analysis using very small changes in L_i. In the case in which the objective of the firm is to equalize the marginal default values across lines of business, the formula for the surplus required per dollar of liabilities in line i (s_i) is:

$$s_i = s - (\partial p/\partial s)^{-1}(\partial p/\partial \sigma) \quad [(\sigma_{iL} - \sigma_L^2) - (\sigma_{iV} - \sigma_{LV})]/\sigma,$$

where
- s = the surplus-to-liability ratio of the firm,
- σ = the firm's overall volatility parameter,
- p = the firm's insolvency put per dollar of total liabilities $= P/(L_1 + L_2 + L_3)$,
- $\partial p/\partial s$ = the partial derivative (rate of change) of p with respect to s,
- $\partial p/\partial \sigma$ = the rate of change of p with respect to the firm's overall volatility parameter σ,
- σ_{iL} = covariance parameter between losses in line i and losses of the entire liability portfolio,
- σ_L^2 = volatility parameter for total losses, L,
- σ_{iV} = covariance parameter between losses in line i and losses of the asset portfolio, and
- σ_{LV} = covariance parameter between the firm's assets and losses.

Thus, because $\partial p/\partial s < 0$ and $\partial p/\partial \sigma > 0$, line i's capital allocation is directly proportional to its covariability with the loss portfolio (σ_{iL}) and inversely proportional to its covariability with the asset portfolio (σ_{iV}). Adding to the covariability of the loss portfolio increases the firm's overall risk level and therefore leads to more capital being allocated to line i. However, because the firm's overall volatility parameter is inversely related to the covariability between assets and liabilities, lines with higher covariability with assets require less equity capital. Intuitively, positive correlation between assets and liabilities creates a natural hedge that reduces the risk of the firm.

The capital allocation for the three-line firm discussed above was calculated using the M-R approach. The results are shown in the last column of Table 3, headed "Myers-Read Marginal Capital." The M-R method allocates 100 percent of total capital. Proportionately to the capital allocated by the two methods, the M-R method allocates less to line 3 and more to lines 1 and 2.[6]

Because the amounts of capital allocated to each line of business differ *substantially* between the M-R and M-P methods, the two methods will not yield the same pricing and project decisions when used with a method that employs by-line capital allocations. Consequently, it is important to determine which method is correct. The M-R method has considerable appeal because it avoids the problem of how to deal with the unallocated capital under the M-P approach. In addition, most decision making regarding pricing and underwriting is marginal in the sense of M-R; i.e., decision making typically involves small changes to an existing portfolio. A possible resolution of the differences between the two methods is that the M-P method is appropriate when adding entire

[6] To calculate this, divide the amount allocated to each line by the total amount allocated: $1,084 for M-P and $1,427 for M-R.

businesses to the firm, whereas the M-R method is appropriate for the firm's normal operations. However, more research is needed to determine which model is more consistent with value maximization.

Additional Issues

Another important issue has to do with the *economic cost* of the firm's overall capital as well as the capital allocated to individual lines of business. The capital of financial institutions such as insurers is invested in marketable securities. If capital markets are efficient and frictionless, the invested funds will earn the equilibrium market rate of return and thus will be cost-less to the firm. However, the existence of various market and institutional imperfections leads to friction costs that imply that the capital invested will not earn the full fair market return required to avoid a loss to the insurer. Various friction costs create costs for insurers that reduce the returns from the investment of their capital. The three most important sources of costly capital to insurers are as follows:

1. Agency and informational costs. Managers of firms may behave opportunistically and thus fail to realize the owners' objective of value maximization. In addition, adverse selection and moral hazard are endemic to insurance markets and will create costs to the extent they cannot be controlled through an insurer's pricing and underwriting decisions.
2. The federal income tax system leads to double taxation of investment income. As a result, investing in securities through an insurance company produces lower investment returns than investors could realize by investing directly in the market.
3. Regulation, and especially the risk-based capital system, imposes costs on insurers in the form of a regulatory "option" on the insurer's assets. The option is created because the RBC system gives regulators the legal right to seize control of the insurer when its assets still exceed its liabilities. Other regulations such as investment restrictions also may lead insurers to hold inefficient portfolios, further reducing returns for any given level of risk.

The existence of market frictions means that a spread develops between the returns that could be earned by investing directly in capital markets and the returns actually earned on the capital held by insurers. This spread cost must be taken into account in determining whether lines of business are earning the appropriate rates of return. Of course, the risk of individual lines is also important in determining their cost of capital. Usually, the type of risk recognized in the cost of capital context is systematic market risk, determined by an asset pricing model.[7]

It is also interesting to consider the role of regulatory risk-based capital in the context of a marginal allocation system. With a well-designed marginal capital-allocation system, is regulatory risk-based capital relevant? It can be, because there is a potential cost imposed on the firm by the regulatory risk-based capital system. For most insurers, and reasonably small EPD targets, no cost will be realized, since the insurer's total capital generally will be greater than the regulatory capital requirement. The author also considers the following two cases:

[7] For further discussion of cost of capital estimation in insurance, see Cummins and Harrington (1987), Cummins and Lamm-Tennant (1994), and Lee and Cummins (1998).

1. Regulatory capital for one or more lines exceeds the marginal capital allocated to these lines but the insurer's overall capital is greater than risk-based capital. In this case, regulatory costs are not likely to be incurred because the risk-based capital test is applied to the entire firm and not by line of business. This conclusion would have to be modified if risk-based capital action level tests were to be applied by line.
2. Regulatory risk-based capital exceeds the firm's overall capital, including both allocated and unallocated capital. In this situation, regulatory penalties will apply.

Thus, the conclusion is that regulatory capital usually will not be a problem, even if one or more lines of business have allocated capital that is less than the by-line risk-based capital, as long as the firm's overall capital exceeds its overall risk-based capital.

The final point to be made is a caveat. Insurers should use caution in designing and implementing capital-allocation systems. The use of an inappropriate system is likely to lead to rejection of some projects that should be accepted and acceptance of some projects that should be rejected. Systems that use stand-alone capital are likely to be particularly harmful. However, the adoption of an appropriate marginal capital-allocation system can have significantly beneficial effects on the market value of the firm by enabling the firm to identify projects and businesses that are creating value for shareholders as well as those that are eroding firm value.

CONCLUSION

It is possible to draw some conclusions about the methodologies that have been discussed in this article. The first is that the expected policyholder deficit provides more information than value at risk. Consequently, if one of these methods is used, the focus should be on the EPD, although it may also be useful to calculate *VaR*. Additionally, exceedence probability and EPD graphs convey much more information than looking at a single *VaR* or EPD value.

Second, using an option model that recognizes diversification will lead to better decisions than *VaR*- or EPD-based methodologies because diversification is an important benefit from operating a multiple-line firm. Between the two approaches that recognize diversification, the Myers-Read method seems to be more effective for the firm's normal operations because most pricing and underwriting decisions are marginal in the sense considered in their methodology. However, more research is clearly needed to determine whether the Myers-Read model actually is more consistent with value-maximization than competing models. A third conclusion is that the cost of capital allocated to a line is the "spread cost" (i.e., the cost over and above the cost of capital if this capital were not held in an insurance company but invested directly in the capital market).

Fourth, capital allocation must consider both asset and liability risk and allow for covariability between assets and liabilities. Capital allocations are increasing in the covariability between a line's losses and the company's overall losses and decreasing in the covariability between a line's losses and the company's asset portfolio. Fifth, the allocation of capital should consider the duration and maturity of liabilities. Even if allocation is done in terms of the liability line lending money to the asset department, the liability line still needs to be charged based on its duration. It is necessary to work out transfer pricing schemes. In this regard, it might be useful to study the public utility literature, where transfer-pricing models are discussed.

Sixth, the decision-making system should drive the design of the data system, and not the other way around. Inevitably, when one starts out, the data will drive the system; but when data systems are redesigned, one needs to think about what kind of information is needed to use capital allocation models. Annual data are insufficient to obtain accurate capital allocations. Therefore, it is critical to obtain data as frequently as possible. Finally, the winning firms in the twenty-first century will be the ones that successfully implement capital allocation and other financial decision-making techniques. Such firms will make better pricing, underwriting, and entry / exit decisions and create value for shareholders.

References

Albrecht, Peter, 1997, Risk Based Capital Allocation and Risk Adjusted Performance Management in Property / Liability Insurance: A Risk Theoretical Framework, *Proceedings of the 75th International AFIR Colloquium* (Cairns, Australia).

Butsic, Robert, 1994, Solvency Measurement for Property-Liability Risk-Based Capital Applications, *Journal of Risk and Insurance,* 61: 656-690.

Cummins, J. David, and Scott E. Harrington, 1988, The Relationship Between Risk and Return: Evidence From Property-Liability Insurance Stocks, *Journal of Risk and Insurance,* 55: 15-31.

Cummins, J. David, Scott E. Harrington, and Robert W. Klein, 1995, Insolvency Experience, Risk-Based Capital, and Prompt Corrective Action in Property-Liability Insurance, *Journal of Banking and Finance,* 19: 511-527.

Cummins, J. David, Scott E. Harrington, and Greg Niehaus, 1993, An Economic Overview of Risk-Based Capital Requirements in the Property-Liability Insurance Industry, *Journal of Insurance Regulation,* 11: 427-447.

Cummins, J. David, and Joan Lamm-Tennant, 1994, Capital Structure and the Cost of Equity Capital in Property-Liability Insurance, *Insurance: Mathematics and Economics,* 15: 187-201.

Fama, Eugene F., and Kenneth R. French, 1996, The CAPM Is Wanted, Dead or Alive, *Journal of Finance,* 51: 1947-1958.

Kaplan, Paul D., and James D. Peterson, 1998, Full-Information Industry Beta, *Financial Management,* 27: 85-93.

Lee, Alice C., and J. David Cummins, 1998, Alternative Models for Estimating the Cost of Equity Capital for Property / Casualty Insurers, *Review of Quantitative Finance and Accounting,* 10: 235-267.

Matten, Chris, 1996, *Managing Bank Capital: Capital Allocation and Performance Measurement* (New York: John Wiley).

Merton, Robert, and Andre Perold, 1993, Theory of Risk Capital in Financial Firms, *Journal of Applied Corporate Finance,* 6: 16-32

Myers, Stewart C., and James A. Read, Jr., 1999, Surplus Allocation for Insurance Companies, working paper (Massachusetts Institute of Technology, Cambridge).

Zaik, Edward, John Walter, Gabriela Kelling, and Christopher James, 1996, RAROC at Bank of America: From Theory to Practice, *Journal of Applied Corporate Finance,* 9: 83-93.

CAS Exam 9 Study Kit

"CAT Bond and Other Risk-Linked Securities: State of Market and Recent Developments"

by J.D. Cummins

© *Risk Management and Insurance Review*, 2008, Vol. 11, No. 1, 23-47

CAT Bonds and Other Risk-Linked Securities: State of the Market and Recent Developments

J. David Cummins

Abstract

This article reviews the current status of the market for catastrophic risk (CAT) bonds and other risk-linked securities. CAT bonds and other risk-linked securities are innovative financial vehicles that have an important role to play in financing mega-catastrophes and other types of losses. The vehicles are especially important because they access capital markets directly, exponentially expanding risk-bearing capacity beyond the limited capital held by insurers and reinsurers. The CAT bond market has been growing steadily, with record amounts of risk capital raised in 2005, 2006, and 2007. CAT bond premia relative to expected losses covered by the bonds have declined by more than one-third since 2001. CAT bonds now appear to be priced competitively with conventional catastrophe reinsurance and comparably rated corporate bonds. CAT bonds have grown to the extent that they now play a major role in completing the market for catastrophic-risk finance and are spreading to other lines such as automobile insurance, life insurance, and annuities. CAT bonds are not expected to replace reinsurance but to complement the reinsurance market by providing additional risk-bearing capacity. Other innovative financing mechanisms such as risk swaps, industry loss warranties, and sidecars also are expected to continue to play an important role in financing catastrophic risk.

Introduction

This article analyzes risk-linked securities as sources of risk capital for the insurance and reinsurance industries. Risk-linked securities are innovative financing devices that enable insurance risk to be sold in capital markets, raising funds that insurers and reinsurers can use to pay claims arising from mega-catastrophes and other loss events. The most prominent type of risk-linked security is the catastrophic risk (CAT) bond, which is a fully collateralized instrument that pays off on the occurrence of a defined catastrophic

---

J. David Cummins is Joseph E. Boettner Professor at Temple University and Harry J. Loman Professor Emeritus, The Wharton School, University of Pennsylvania, 1301 Cecil B. Moore Avenue, 481 Ritter Annex, Philadelphia, PA 19122; phone: 215-204-8468, 610-520-9792; fax: 610-520-9790; e-mail: cummins@temple.edu. The author thanks Roger Beckwith, William Dubinsky, Morton Lane, and Christopher M. Lewis for helpful comments. Any errors or omissions are the responsibility of the author.

event. CAT bonds and other risk-linked securities are potentially quite important because they have the ability to access the capital markets to provide capacity for insurance and reinsurance markets. The CAT bond market has expanded significantly in recent years and now seems to have reached critical mass. Although the CAT bond market is small in comparison with the overall nonlife reinsurance market, it is of significant size in comparison with the property-catastrophe reinsurance market. Some industry experts observe that nontraditional risk financing instruments, including CAT bonds, industry loss warranties (ILWs), and sidecars, now represent the majority of the property-catastrophe retrocession market.

This article begins by discussing the design of CAT bonds and other risk-linked securities. The discussion then turns to the evolution of the risk-linked securities market and an evaluation of the current state of the market. The scope of the article is limited primarily to securitization of catastrophic property-casualty risks. However, there also are rapidly developing markets in automobile and other types of noncatastrophe insurance securitizations as well as life insurance securitizations, which are discussed in Cowley and Cummins (2005).

THE STRUCTURE OF RISK-LINKED SECURITIES

This section considers the structure of CAT bonds and other risk-linked securities that have been used to raise risk capital for property-casualty risks. The discussion focuses primarily on CAT bonds but also considers other innovative risk financing solutions. Included in the latter category are some investment structures that are not necessarily securities in the sense of being tradable financial instruments but are innovative approaches whereby insurers and reinsurers can either access capital markets to supplement traditional reinsurance.

Risk-Linked Securities: Early Developments

Following Hurricane Andrew in 1992, efforts began to access securities markets directly as a mechanism for financing future catastrophic events. The first contracts were launched by the Chicago Board of Trade (CBOT), which introduced catastrophe futures in 1992 and later introduced catastrophe put and call options. The options were based on aggregate catastrophe loss indices compiled by Property Claims Services (PCS), an insurance industry statistical agent.[1] The contracts were later withdrawn due to lack of trading volume. In 1997, the Bermuda Commodities Exchange (BCE) also attempted to develop a market in catastrophe options, but the contracts were withdrawn within 2 years as a result of lack of trading.

Insurers had little interest in the CBOT and BCE contracts for various reasons, including the thinness of the market, possible counterparty risk on the occurrence of a major catastrophe, and the potential for disrupting long-term relationships with reinsurers. Another concern with the option contracts was the possibility of excessive basis risk, i.e., the risk that payoffs under the contracts would be insufficiently correlated with insurer losses. A study by Cummins et al. (2004) confirms that basis risk was a legitimate concern.

[1] Contracts were available based on a national index, five regional indices, and three state indices, for California, Florida, and Texas. For further discussion, see Cummins (2005).

Interestingly, in 2007 two separate exchanges, the Chicago Mercantile Exchange (CME) and the New York Mercantile Exchange (NYMEX) introduced futures and options contracts on U.S. hurricane risk. Both exchanges indicate in their distributional materials on the contracts that their introduction was motivated by the 2005 U.S. hurricane season, which revealed the limitations on the capacity of insurance and reinsurance markets. CME currently lists contracts on hurricanes in six U.S. regions: the Gulf Coast, Florida, Southern Atlantic Coast, Northern Atlantic Coast, Eastern United States, and Galveston-Mobile. CME contracts settle on the Carvill Hurricane Indices created by Carvill, a reinsurance intermediary. NYMEX initial listings were a U.S. national contract, a Florida contract, and a Texas-to-Maine contract. The NYMEX contracts will settle on catastrophe loss indices. The NYMEX indices are calculated by Gallagher Re based on data provided by Property Claims Services, the same data source utilized for the earlier CBOT options. Given that both the CME and NYMEX contracts are based on broadly defined geographical areas, they will be subject to significant basis risk. Thus, it remains to be seen whether these contracts will succeed where the similar CBOT contracts failed. However, given the existence of a secondary market as well as dedicated CAT bond mutual funds, it is possible that the CME or NYMEX contracts could be used for hedging purposes by investors with broadly diversified portfolios of CAT bonds.

Another early attempt at securitization involved contingent notes known as "Act of God" bonds. In 1995, Nationwide issued $400 million in contingent notes through a special trust—Nationwide Contingent Surplus Note (CSN) Trust. Proceeds from the sale of the bonds were invested in 10-year Treasury securities, and investors were provided with a coupon payment equal to 220 basis points over Treasuries. Embedded in these contingent capital notes was a "substitutability" option for Nationwide. Given a prespecified event that depleted Nationwide's equity capital, Nationwide could substitute up to $400 million of surplus notes for the Treasuries in the Trust at any time during a 10-year period for any "business reason," with the surplus notes carrying a coupon of 9.22 percent.[2] Although two other insurers issued similar notes, this type of structure did not achieve a significant segregation of Nationwide's liabilities, leaving investors exposed to the general business risk of the insurer and to the risk that Nationwide might default on the notes. In addition, unlike CAT bonds, the withdrawal of funds from the trust would create the obligation for Nationwide eventually to repay the Trust. Consequently, contingent notes have not emerged as a major solution to the risk-financing problem.

CAT Bonds

The securitized structure that has achieved the greatest degree of success is the CAT bond. CAT bonds were modeled on asset-backed-security transactions that have been executed for a wide variety of financial assets including mortgage loans, automobile loans, aircraft leases, and student loans. CAT bonds are part of a broader class of assets known as *event-linked bonds*, which pay off on the occurrence of a specified event. Most event-linked bonds issued to date have been linked to catastrophes such as hurricanes and earthquakes, although bonds also have been issued that respond to mortality events.

[2] Surplus notes are debt securities issued by mutual insurance companies that regulators treat as equity capital for statutory accounting purposes. The issuance of such notes requires regulatory approval.

The first successful CAT bond was an $85 million issue by Hannover Re in 1994 (Swiss Re, 2001). The first CAT bond issued by a nonfinancial firm, occurring in 1999, covered earthquake losses in the Tokyo region for Oriental Land Company, the owner of Tokyo Disneyland. Although various design features were tested in the early stages of the CAT bond market, more recently CAT bonds have become more standardized. The standardization has been driven by the need for bonds to respond to the requirements of the principal stakeholders including sponsors, investors, rating agencies, and regulators.

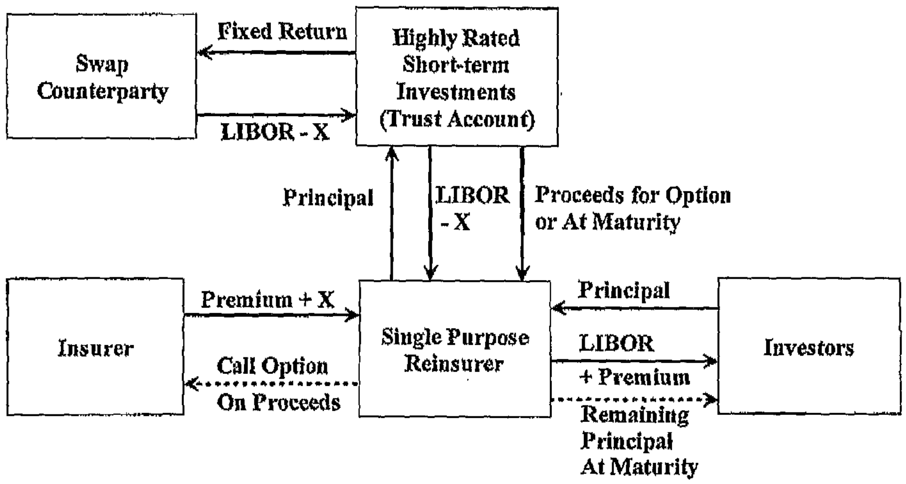
CAT bonds often are issued to cover the so-called *high layers* of reinsurance protection, e.g., protection against events that have a probability of occurrence of 0.01 or less (i.e., a *return period* of at least 100 years). The higher layers of protection often go unreinsured by ceding companies for two primary reasons—for events of this magnitude, ceding insurers are more concerned about the credit risk of the reinsurer, and high layers tend to have the highest reinsurance margins or pricing spreads above the expected loss (Cummins, 2007). Because CAT bonds are fully collateralized, they eliminate concerns about credit risk, and because catastrophic events have low correlations with investment returns, CAT bonds may provide lower spreads than high-layer reinsurance because they are attractive to investors for diversification.

CAT bonds also can lock in multi-year protection, unlike traditional reinsurance, which usually is for a 1-year period, and shelter the sponsor from cyclical price fluctuations in the reinsurance market. The multi-year terms (or *tenors*) of most CAT bonds also allow sponsors to spread the fixed costs of issuing the bonds over a multi-year period, reducing costs on an annualized basis.

A typical CAT bond structure is diagrammed in Figure 1. The transaction begins with the formation of a single purpose reinsurer (SPR). The SPR issues bonds to investors and invests the proceeds in safe, short-term securities such as government bonds or AAA corporates, which are held in a trust account. Embedded in the bonds is a call option that is triggered by a defined catastrophic event. On the occurrence of the event, proceeds are released from the SPR to help the insurer pay claims arising from the event. In most CAT bonds, the principal is fully at risk, i.e., if the contingent event is sufficiently large, the investors could lose the entire principal in the SPR. In return for the option, the insurer pays a premium to the investors. The fixed returns on the securities held in the trust are usually swapped for floating returns based on London interbank offered rate (LIBOR) or some other widely accepted index. The reason for the swap is to immunize the insurer and the investors from interest rate (mark-to-market) risk and also default risk. The investors receive LIBOR plus the risk premium in return for providing capital to the trust. If no contingent event occurs during the term of the bonds, the principal is returned to the investors upon the expiration of the bonds.

Some CAT bond issues have included *principal protected tranches*, where the return of principal is guaranteed. In this tranche, the triggering event would affect the interest and spread payments and the timing of the repayment of principal. For example, a 2-year CAT bond subject to the payment of interest and a spread premium might convert into a 10-year zero-coupon bond that would return only the principal. Principal-protected tranches have become relatively rare, primarily because they do not provide as much risk capital to the sponsor as a principal-at-risk bond.

FIGURE 1
CAT Bond With Single-Purpose Reinsurer



Insurers prefer to use a SPR to capture the tax and accounting benefits associated with traditional reinsurance.[3] Investors prefer SPRs to isolate the risk of their investment from the general business and insolvency risks of the insurer, thus creating an investment that is a "pure play" in catastrophic risk. In addition, the bonds are fully collateralized, with the collateral held in trust; insulating the investors from credit risk. As a result, the issuer of the securitization can realize lower financing costs through segregation. The transaction also is more transparent than a debt issue by the insurer, because the funds are held in trust and are released according to carefully defined criteria.

The bonds are attractive to investors because catastrophic events have low correlations with returns from securities markets and hence are valuable for diversification purposes (Litzenberger et al., 1996). Although the $100 billion-plus "Big One" hurricane or earthquake could drive down securities prices, creating systematic risk for CAT securities, systematic risk is considerably lower than for most other types of assets, especially during more normal periods.

In the absence of a traded underlying asset, CAT bonds and other insurance-linked securities have been structured to pay off on three types of triggering variables: (1) *indemnity triggers*, where payouts are based on the size of the sponsoring insurer's actual losses; (2) *index triggers*, where payouts are based on an index not directly tied to the sponsoring firm's losses; or (3) *hybrid triggers*, which blend more than one trigger in a single bond.

There are three broad types indices that can be used as CAT bond triggers—industry loss indices, modeled loss indices, and parametric indices. With industry loss indices, the

[3] Harrington and Niehaus (2003) argue that one important advantage of CAT bonds as a financing mechanism is that corporate tax costs are lower than for financing through equity and that the bond poses less risk in terms of potential future degradations of insurer financial ratings and capital structure than financing through subordinated debt.

payoff on the bond is triggered when estimated industry-wide losses from an event exceed a specified threshold. For example, the payoff could be based on estimated catastrophe losses in a specified geographical area provided by Property Claims Services (PCS), the same organization that provided the indices for the CBOT options. A modeled-loss index is calculated using a model provided by one of the major catastrophe-modeling firms—Applied Insurance Research Worldwide, EQECAT, or Risk Management Solutions. The index could be generated by running the model on industry-wide exposures for a specified geographical area. Alternatively, the model could be run on a representative sample of the sponsoring insurer's own exposures. In each case, an actual event's physical parameters are used in running the simulations. Finally, with a parametric trigger, the bond payoff is triggered by specified physical measures of the catastrophic event such as the wind speed and location of a hurricane or the magnitude and location of an earthquake.

There are a number of factors to consider in the choice of a trigger when designing a CAT bond (Guy Carpenter, 2005a; Mocklow et al., 2002). The choice of a trigger involves a trade-off between moral hazard; (transparency to investors) and basis risk. Indemnity triggers are often favored by insurers and reinsurers because they minimize basis risk, i.e., the risk that the loss payout of the bond will be greater or less than the sponsoring firm's actual losses. However, indemnity triggers require investors to obtain information on the risk exposure of the sponsor's underwriting portfolio. This can be difficult, especially for complex commercial risks. In addition, indemnity triggers have the disadvantage to the sponsor that they require disclosure of confidential information on the sponsor's policy portfolio. Contracts based on indemnity triggers may require more time than nonindemnity triggers to reach final settlement because of the length of the loss adjustment process.

Index triggers tend to be favored by investors because they minimize the problem of moral hazard; i.e., they maximize the transparency of the transaction. Moral hazard can occur if the issuing insurer fails to settle catastrophe losses carefully and appropriately (i.e., overpays) because of the correlation of the bond payout with its realized losses. The insurer might also excessively expand its premium writings in geographical areas covered by the bond. Although CAT bonds almost always contain copayment provisions to control moral hazard, moral hazard remains a residual concern for some investors. Indices also have the advantage of being measurable more quickly after the event than indemnity triggers, so that the sponsor receives payment under the bond more quickly.

The principal disadvantage of index triggers is that they expose the sponsor to a higher degree of basis risk than indemnity triggers. The degree of basis risk varies depending upon several factors. Parametric triggers tend to have the lowest exposure to moral hazard but may have the highest exposure to basis risk. However, even with a parametric trigger, basis risk can be often be reduced substantially by appropriately defining the location where the event severity is measured. Similarly, industry loss indices based on narrowly defined geographical areas tend to have less basis risk than those based on wider areas (Cummins et al., 2004). Modeled-loss indices may become the favored mechanism for obtaining the benefits of an index trigger without incurring significant basis risk. However, modeled-loss indices are subject to "model risk," i.e., the risk that the model will over- or underestimate the losses from an event. This risk is diminishing over time as the modeling firms continue to refine their models.

Sidecars

An innovative financing vehicle with some similarities to both conventional reinsurance and CAT bonds is the *sidecar.* Sidecars date back to at least 2002 but became much more prominent following the 2005 hurricane season (A.M. Best Company, 2006). Sidecars are special purpose vehicles formed by insurance and reinsurance companies to provide additional capacity to write reinsurance, usually for property catastrophes and marine risks, and typically serve to accept retrocessions exclusively from a single reinsurer. Sidecars are typically off-balance sheet, formed to write specific types of reinsurance such as property-catastrophe quota share or excess of loss, and generally have limited lifetimes. Sidecars and excess of loss CAT bonds can work together as complementary instruments in much the same way as quota share and excess of loss complement each other in a traditional reinsurance program.

Reinsurers receive override commissions for premiums ceded to sidecars. Most sidecars are capitalized by private investors such as hedge funds, but insurers and reinsurers also participate in this financing device. Sidecars receive premiums for the reinsurance underwritten and are liable to pay claims under the terms of the reinsurance contracts. In addition to providing capacity, sidecars also enable the sponsoring reinsurer to move some of its risks off-balance sheet, thus improving leverage. Sidecars can also be formed quickly and with minimal documentation and administrative costs.[4]

Catastrophic Equity Puts (Cat-E-Puts)

Another capital market solution to the catastrophic loss financing problem is catastrophic equity puts (Cat-E-Puts). Unlike CAT bonds, Cat-E-Puts are not asset-backed securities but options. In return for a premium paid to the writer of the option, the insurer obtains the option to issue preferred stock at a preagreed price on the occurrence of a contingent event. This enables the insurer to raise equity capital at a favorable price after a catastrophe, when its stock price is likely to be depressed. Cat-E-Puts tend to have lower transactions costs than CAT bonds because there is no need to set up an SPR. However, because they are not collateralized, these securities expose the insurer to counterparty performance risk. In addition, issuing the preferred stock can dilute the value of the firm's existing shares. Thus, although Cat-E-Puts have been issued, they have not become nearly as important as CAT bonds.

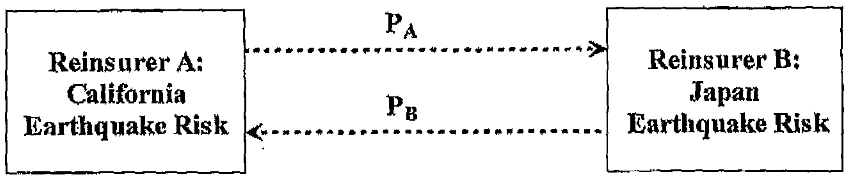
Catastrophe Risk Swaps

Like Cat-E-Puts, catastrophe risk swaps generally are not prefunded but rely only an agreement between two counterparties. Catastrophe swaps can be executed between two firms with exposure to different types of catastrophic risk. An example of a catastrophic-risk swap is provided in Figure 2. In the example, a reinsurer with exposure to California earthquake risk agrees to swap its risk with another reinsurer with exposure to Japanese earthquake risk. Another example is the swap executed by Mitsui Sumitomo Insurance and Swiss Re in 2003, which swapped $50 million of Japanese typhoon risk against $50 million of North Atlantic hurricane risk and $50 million of Japanese typhoon risk against $50 million of European windstorm risk. In some instances, a reinsurer may serve as an intermediary between the swap partners, but in most instances CAT swaps

[4] For further discussion, see Cummins (2007) and Lane (2007).

FIGURE 2
Catastrophe Risk Swap

Reinsurers A and B Swap Earthquake Risk



P_A = contingent payment for Japan earthquake

P_B = contingent payment for California earthquake

are done directly between two (re)insurers. Swaps are facilitated by the Catastrophic Risk Exchange (CATEX), a web-based exchange where insurers and reinsurers can arrange reinsurance contracts and swap transactions.

The event or events that trigger payment under the swap are carefully defined in the swap agreement. For example, a parametric trigger could be used such as an earthquake of a specified magnitude in Tokyo for the Japanese side of the swap and a comparable earthquake in San Francisco for the U.S. side. The swap can be designed such that the two sides of the risk achieve *parity*, i.e., such that the expected losses under the two sides of the swap are equivalent. This obviously requires an extensive modeling exercise, which would be conducted using one of the models developed by catastrophe-modeling firms or internally. With parity, there is no exchange of money at the inception of the contract, only on the occurrence of one of the triggering events. The swap also defines a specified amount of money to be paid if an event occurs, such as $200 million. Some contracts have sliding scale payoff functions, which specify full payout for the severest events and partial payout for smaller events. Swaps can be annual or can span several years. Swaps also can be executed that fund multiple risks simultaneously such as also swapping North Atlantic hurricane risk for Japanese typhoon risk in the same contract as the earthquake swap.

Swaps may be attractive substitutes for reinsurance, CAT bonds, and other risk financing devices. They have the advantage that the reinsurer simultaneously lays of some of its core risk and obtains a new source of diversification by exchanging uncorrelated risks with the counterparty (Takeda, 2002). Thus, swaps may enable reinsurers to operate with less equity capital. Swaps also are characterized by low transactions costs and reduce current expenses because no money changes hands until the occurrence of a triggering event. The potential disadvantages of swaps are that modeling the risks to achieve parity can be challenging and is not necessarily completely accurate. Swaps also may create more exposure to basis risk than some other types of contracts and also create exposure to counter-party nonperformance risk. The possibility of nonperformance risk provides another potential role for an investment bank or specialized reinsurer to execute hedges to enhance the credit quality of the swap. However, such hedging would add to the transactions costs of the deal. Systematic data on the magnitude of the risk swaps market presently are not available. However, industry experts interviewed by the author indicate that the swaps market is "quite substantial."

ILW

As explained further below, a possible impediment to the growth of the CAT securitization market has to do with whether the securities are treated as reinsurance by regulators, and hence given favorable regulatory accounting treatment. It seems clear that properly structured *indemnity* CAT securities (those that pay off based on the losses of the issuing insurer) will be treated as reinsurance. Nevertheless, regulation does not seem to have impeded the strong growth of the CAT bond market during the past several years because sponsors and their bankers have found various ways to finesse potential regulatory problems. For example, even if the SPV is an offshore vehicle, the trust holding the assets can be onshore, mitigating regulatory concerns regarding credit risk of offshore entities.

Dual-trigger contracts known as *industry loss warranties* (ILW) also overcome regulatory objections to nonindemnity bonds (McDonnell, 2002). ILWs are dual-trigger reinsurance contracts that have a *retention trigger* based on the incurred losses of the insurer buying the contract and also a *warranty trigger* based on an industry-wide loss index. That is, the contracts pay off on the dual event that a specified industry-wide loss index exceeds a particular threshold at the same time that the issuing insurer's losses from the event equal or exceed a specified amount. Both triggers have to be hit in order for the buyer of the contract to receive a payoff. The issuing insurer thus is covered in states of the world when its own losses are high and the reinsurance market is likely to enter a hard-market phase. ILWs cover events from specified catastrophe perils in a defined geographical region. For example, an ILW might cover losses from hurricanes in the Southeastern United States. The term of the contact is typically 1 year. ILWs may have *binary triggers*, where the full amount of the contract pays off once the two triggers are satisfied or *pro rata triggers* where the payoff depends upon how much the loss exceeds the warranty.

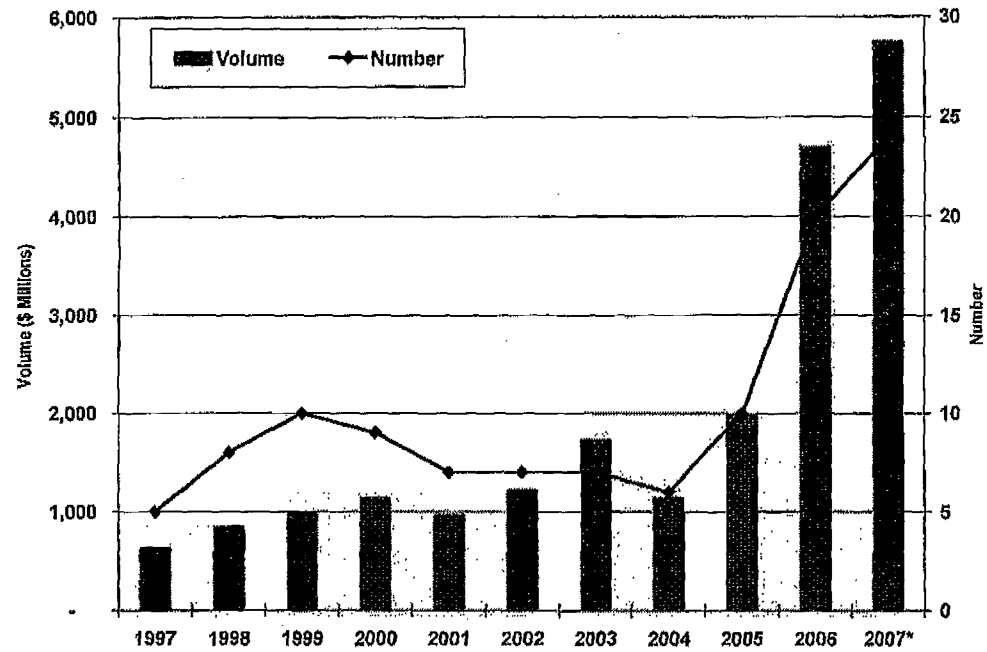
The principal advantages of ILWs are that they are treated as reinsurance for regulatory purposes, and that they can be used to plug gaps in reinsurance programs. They also represent an efficient use of funds in that they pay off in states of the world where both the insurer's losses and industry-wide losses are high.

Systematic data on the size of the ILW market are presently not available. However, reinsurance experts interviewed by the author believe that the ILW market is roughly of the same order of magnitude as the CAT bond market. Experts also comment that capital market participants provide the majority of risk capital in the ILW market, just as they do in the CAT bond market. ILWs can be packaged and securitized, broadening the investor base.

THE RISK-LINKED SECURITIES MARKET

This section reviews the recent history and current status of the risk-linked securities market. The focus is primarily on CAT bonds, which are the most commonly used securitized structure used in financing catastrophic risk.

FIGURE 3
Nonlife CAT Bonds: New Issues



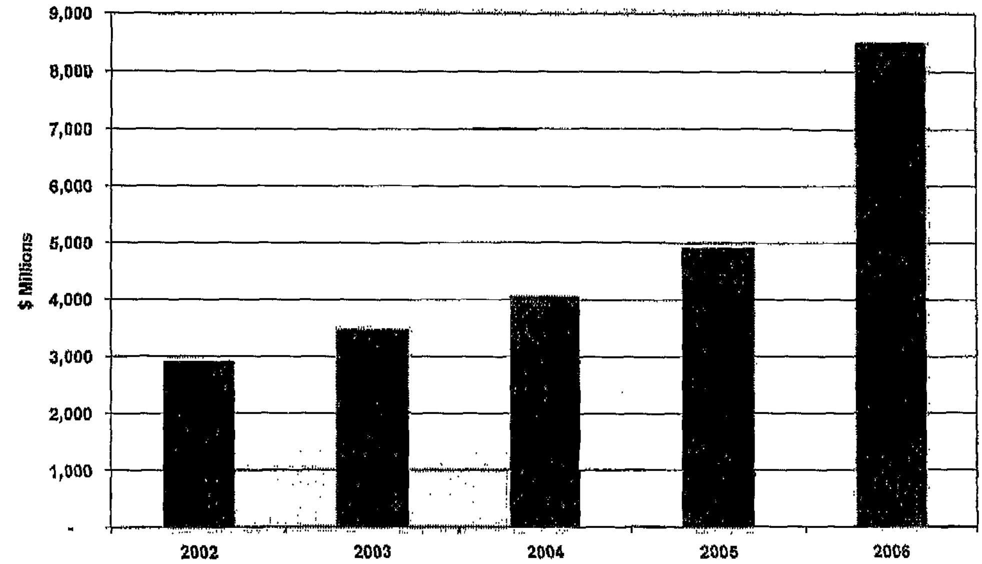
*Through July 31, 2007.
Source: MMC Securities (2007) and Swiss Re (2007b).

The CAT Bond Market: Size and Bond Characteristics

Although the CAT bond market seemed to get off to a slow start in the late 1990s, the market has matured and now has become a steady source of capacity for both primary insurers and reinsurers. The market is growing steadily and set new records for market issuance volume in 2005, 2006, and 2007. CAT bonds make sound economic sense as a mechanism for funding mega-catastrophes. Catastrophes such as Hurricane Katrina and the fabled and yet to be realized $100 billion-plus "Big One" in California, Tokyo, or Florida are large relative to the resources of the insurance and reinsurance industries but are small relative to the size of capital markets (Cummins, 2006). A $100 billion loss would represent less than 0.5 of 1 percent of the value of U.S. securities markets and could easily be absorbed through securitized transactions. Securities markets also are more efficient than insurance markets in reducing information asymmetries and facilitating price discovery. Thus, it makes sense to predict that the CAT bond market will continue to grow and that CAT bonds will eventually be issued in the public securities markets, rather than being confined primarily to private placements as at present.

The new issue volume in the CAT bond market from 1997 through July 2007 is shown in Figure 3. The data in the figure apply only to nonlife CAT bonds. Recently, event-linked bonds have also been issued to cover third-party commercial liability, automobile quota share, and indemnity-based trade credit reinsurance. There is also a growing market in life insurance securitizations of various types.

FIGURE 4
CAT Bonds: Risk Capital Outstanding



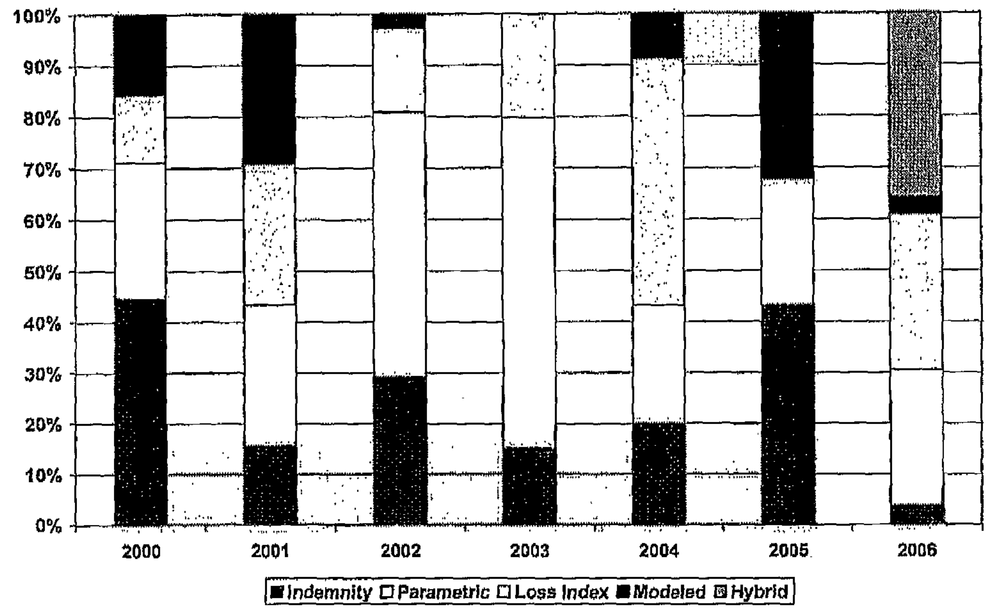
Source: Guy Carpenter (2006a) and MMC Securities (2007).

Figure 3 shows that the market has grown from less than $1 billion per year in 1997 to more than $2 billion per year in the first half of 2005, and then accelerated to nearly $5 billion in 2006 and nearly $6 billion in the first 7 months of 2007. The number of transactions also has been increasing, to 24 in the first 7 months of 2007. A substantial number of the issuers in 2005–2007 were first-time sponsors of CAT bonds, although established players such as Swiss Re continue to play a major role (Guy Carpenter, 2007). Figure 4 shows that the amount of risk capital outstanding in CAT bond markets has also grown steadily. Risk-capital outstanding represents the face value of all bonds still in effect in each year shown in the figure. Nearly $9 billion of risk capital was outstanding by the end of 2006, and nearly $14 billion by mid 2007 (Swiss Re, 2007b).

The characteristics of CAT bonds continue to evolve, but the overall trend is toward a higher degree of standardization. The issue volume by trigger type between 2000 and 2006 is shown in Figure 5. For the period as a whole, index or hybrid bonds accounted for 80 percent of total issue volume. The leading type of index by issue volume is the parametric index, accounting for 34 percent of total issuance. Indemnity bonds made a come-back in 2005 but fell off again in 2006.

The trends in bond tenor are shown in Figure 6. Even though there were some 10-year bonds issued during the 1990s, the market seems to have converged on shorter-term issues, with 3-year bonds constituting the majority of issues in 2005 and 2006. Maturities greater than 1 year tend to be favored because they provide a steady source of risk capital that is insulated from year-to-year swings in reinsurance prices and because they permit issuers to amortize costs of issuance over a longer period, reducing per period transactions costs. Bonds longer than 5 years are not favored by the market

FIGURE 5
CAT Bond Issues by Trigger Type



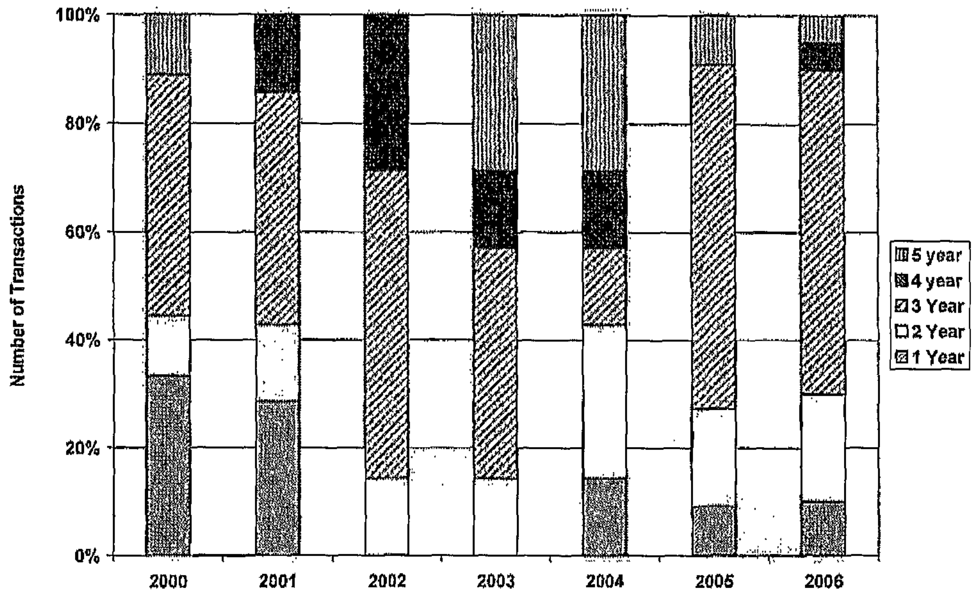
Source: Guy Carpenter (2006a) and MMC Securities (2007).

because market participants would like to reprice the risk periodically to reflect new information on the frequency and severity of catastrophes and to recognize changes in the underwriting risk profile of the sponsor.

For the period as a whole, insurers accounted for 47.9 percent of bonds by issue volume, reinsurers accounted for 47.5 percent, and corporate/government issues accounted for 4.7 percent. In 2006, the first government issued disaster-relief bond placement was executed to provide funds to the government of Mexico to defray costs of disaster recovery.

Specifically, the Mexican bonds would pay off to the benefit of the Mexican Natural Disaster Fund (FONDEN). The CAT bonds are limited to Mexican earthquake risk, but future bonds may be issued that cover Mexican hurricane risk. The bonds were part of a $450 million reinsurance transaction with European Finance Reinsurance, a wholly owned subsidiary of Swiss Re. Swiss Re retained $290 million of the contract exposure and issued $160 million in CAT bonds (notes) with a 3-year bond tenor through a special purpose vehicle, CAT-Mex Ltd. The bonds are binary and parametric, triggered by earthquake physical parameters, including Richter scale readings. Two tranches were issues covering different Mexican earthquake zones. The larger tranche ($150 million) has an expected annual loss of 0.96 percent and a spread over LIBOR of 235 basis points, whereas the smaller tranche ($10 million) has an expected annual loss of 0.93 percent and a spread of 230 basis points. The Mexican bonds provide another indication that the spreads on CAT bonds are declining and show that opportunities exist for securitization

FIGURE 6
CAT Bond Transactions by Bond Tenor



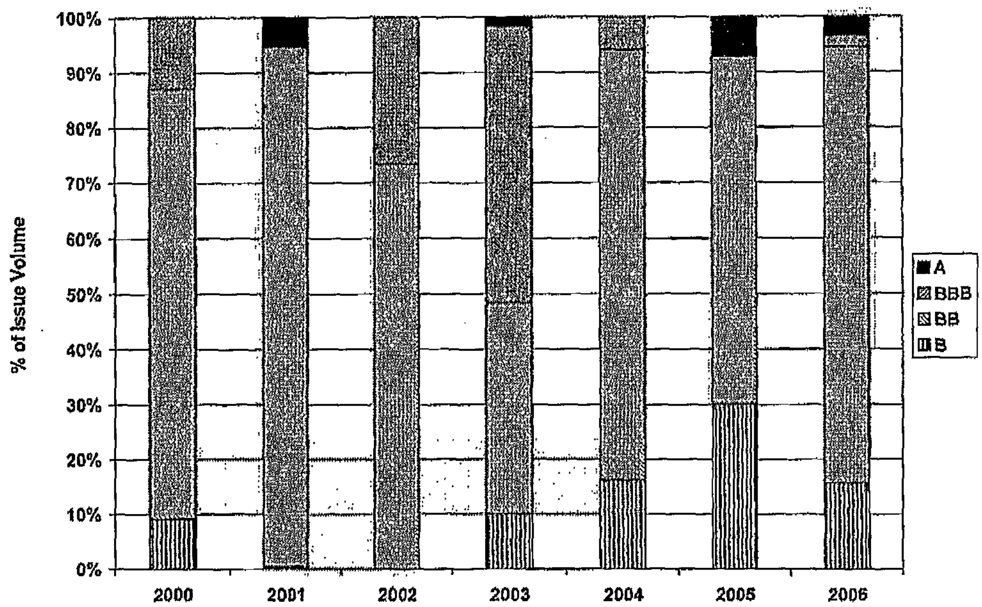
Source: Guy Carpenter (2006a) and MMC Securities (2007).

elsewhere in the world.[5] The bonds are also important because they illustrate how securitization can be used by governments to prefund disaster relief programs, rather than waiting for disaster relief from donor countries and international financial organizations *ex post*.

Obtaining a financial rating is a critical step in issuing a CAT bond because buyers use ratings to compare yields on CAT bonds with other corporate securities. Consequently, almost all bonds are issued with financial ratings. The ratings by bond issue volume from 2000 through 2006 are shown in Figure 7. The vast majority of CAT bonds issued in 2005 and 2006 have been below investment grade (ratings below BBB); i.e., 93 percent of the 2005 issuance volume and 94.5 percent of the 2006 volume were rated BB or B. In 2007, there has been a resurgence in investment-grade bonds (Swiss Re, 2007b), although the majority of CAT bonds are below investment grade in 2007 as well. Although lower than investment grade bond ratings are generally bad news for insurers, reinsurers, and other corporate bond issuers, they are not necessarily adverse in the CAT bond market. Because CAT bonds are fully collateralized, CAT bond ratings tend to be determined by the probability that the bond principal will be hit by a triggering event. Thus, the bond ratings merely indicate the layer of catastrophic-risk coverage that is being provided by the bonds. Although it is important for CAT bonds to be issued with financial ratings, the modeling firm's analysis drives the price more than the actual rating.

[5] For further discussion of the Mexican bonds, see Cardenas et al. (Forthcoming).

FIGURE 7
CAT Bond Issue Volume by Financial Rating

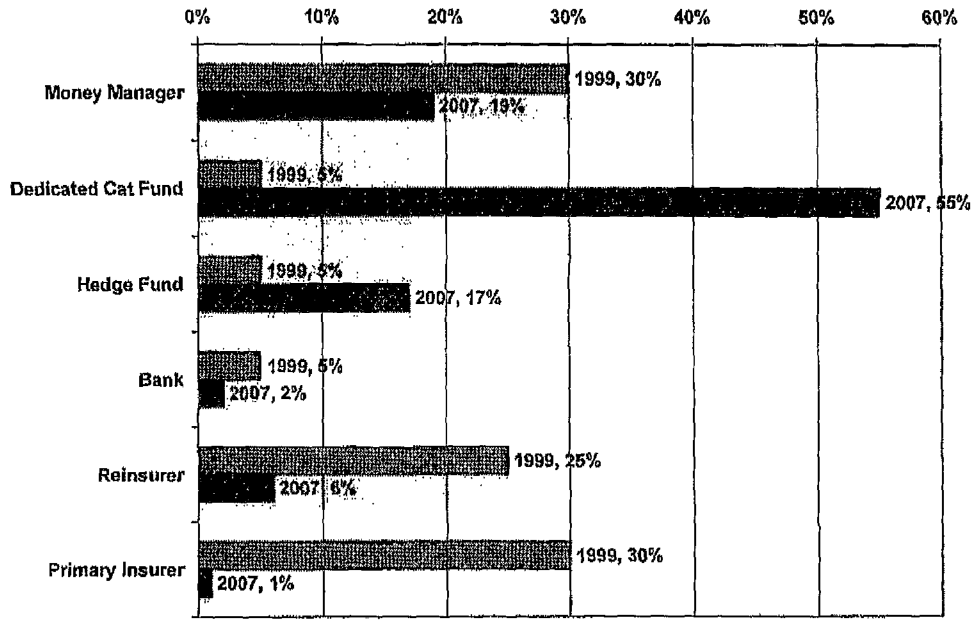


Source: Guy Carpenter (2006a) and MMC Securities (2007).

In the past, the CAT bond market has been criticized for lack of investor interest. However, that critique of the market is now out of date—recent data suggest that there is broad market interest in CAT bonds among institutional investors. Figure 8 shows the percentage of new issue volume by investor type in 1999 and 2007. In 1999, insurers and reinsurers were among the leading investors in the bonds, accounting for 55 percent of the market; i.e., insurers were very prominent on both the supply and demand sides of the market. If insurers and reinsurers are on both sides of the market, the market cannot be said to have attracted very much new capital into the financing of catastrophic risk. However, by 2007, insurers and reinsurers accounted for only 7 percent of demand, suggesting that substantial external capital has been attracted to the market. Dedicated CAT funds accounted for 55 percent of the market in 2007, and money managers and hedge funds accounted for 36 percent. The declining spreads and increasingly broad market interest in the bonds suggest that the bonds are attractive to investors and are playing an increasingly important role relative to conventional reinsurance.

In addition to CAT bonds, a significant amount of new capital was raised through sidecars in 2005 and 2006. The new capital raised through Bermuda sidecars in 2006 is shown in Table 1. Eleven sidecar transactions took place in 2006, totaling $2.9 billion in risk capital. In 2005, there were eight transactions, which raised a total of $2.5 billion. There was some indication that sidecars were competing with CAT bonds for risk capital of interested investors in 2005, leading to rising prices and tightening capacity in the CAT bond market (Guy Carpenter, 2006a). However, the CAT bond market clearly rebounded in 2006 and 2007.

Figure 8
CAT Bonds: New Issue Volume Purchased by Investor Type



Source: Swiss Re.

The first publicly acknowledged total loss of principal for a CAT bond took place in 2005, although there apparently have been earlier wipeouts that were not publicly announced (Lane and Beckwith, 2006). KAMP Re 2005 Ltd., a $190 million bond issued in July 2005 under the sponsorship of Zurich Financial, apparently paid out its entire principal to the sponsor as a result of Hurricane Katrina claims (Guy Carpenter, 2006a). KAMP Re had an indemnity trigger, and the short-term impact of the wipeout was to increase investor wariness of indemnity-based transactions. However, indemnity transactions rebounded in 2007 due to a surge of primary insurer CAT bond issues (Swiss Re, 2007b).

The longer-term impact of the KAMP Re wipeout on the CAT bond market is likely to be favorable. The smooth settlement of the KAMP Re bond established an important precedent in the market, showing that CAT bonds function as designed, with minimal confusion and controversy between the sponsor and investors. Thus, the wipeout served to "reduce the overall uncertainty associated with this marketplace and therefore increase both investor and sponsor demand for these instruments" (Guy Carpenter, 2006a, p. 4).

CAT Bond Prices

CAT bonds are priced at spreads over LIBOR, meaning that investors receive floating interest plus a spread or premium over the floating rate. In the past, CAT bonds have been somewhat notorious for having high spreads, and much has been written trying to explain the magnitude of the spreads (e.g., Froot, 2001). However, there are now significant indications that the spreads are not as high as they might seem relative to the cost of reinsurance, such that CAT bonds are more competitive with conventional reinsurance than earlier analyses may have suggested.

TABLE 1
New Capital Raised Through Sidecars in 2006 (US$ Millions)

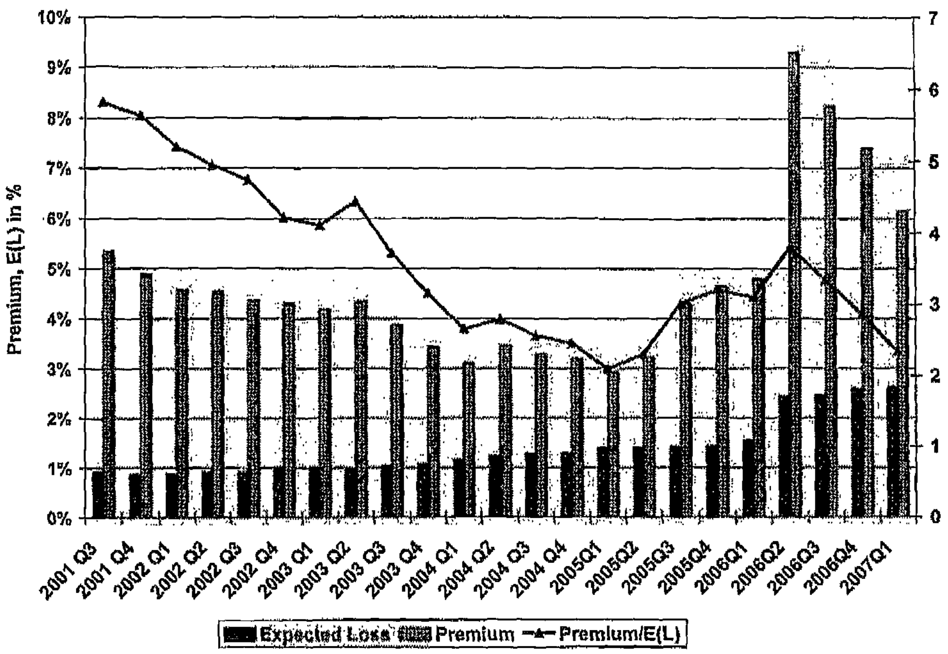
Vehicle Name	Sponsor	Equity	Debt	Total
Bay Point Re	Harborpoint	125	125	250
Concord Re	Lexington Insurance	375	375	750
Helicon Re	White Mountains	145	185	330
Monte Forte Re	Flagstone Re	60		60
Panther Re	Hiscox	144	216	360
Petrel Re	Validus Re	200		200
Sirocco Re	Lancashire Re	95		95
Starbound Re	Rennaissance Re	127	184	311
Stoneheath Re	XL Re	300		300
Timicuan Re	Rennaissance Re	50	20	70
Triomphe Re	Paris Re	121	64	185
Total		1,742	1,169	2,911

Source: MMC Securities (2007).

Because CAT bonds are not publicly traded, it is difficult to obtain data on CAT bond yields. However, there is an active, though nonpublic, secondary market that provides some guidance on yields. The secondary market yields on CAT bonds are shown quarterly from the third quarter of 2001 through the first quarter of 2007 in Figure 9. The numbers in the figure reflect investment yields over LIBOR. The figure shows the absolute yields and also an estimate of the expected loss. The data are from Lane and Beckwith (2005, 2006, 2007a, 2007b). Figure 9 shows the expected loss, the premium, and the bond spread (ratio of premium to expected loss), based on averages of secondary market transactions.

Prior to Katrina, there was a more or less steady decline in yields and a slight increase in the expected loss, implying a general decline in the cost of financing through CAT bonds. The ratio of the premium to expected loss was about six in early 2001, and prior research covering periods before 2001, showed median ratios of yields to expected loss of about 6.5 for CAT bonds (Cummins et al., 2004). However, the ratio of premium to expected loss began a more or less steady decline between 2001 and 2005 and stood at 2.1 in the first quarter of 2005. Not surprisingly, yields and spreads increased following Katrina as the market tightened and investors had opportunities to place capital in other catastrophic-risk vehicles such as sidecars. The spread peaked at 3.7 in the second quarter of 2006 but declined again to 2.3 by the first quarter of 2007. Thus, the CAT bond market was able to withstand the post-Katrina competition for capital without returning to the high relative spreads of earlier periods. Consequently, it seems that the earlier critique of CAT bonds, i.e., excessive spreads, no longer applies. This is the expected result in a market where there is growing investor interest and expertise as well as growing volume, which adds to market liquidity.

FIGURE 9
CAT Bond Premia and Expected Loss



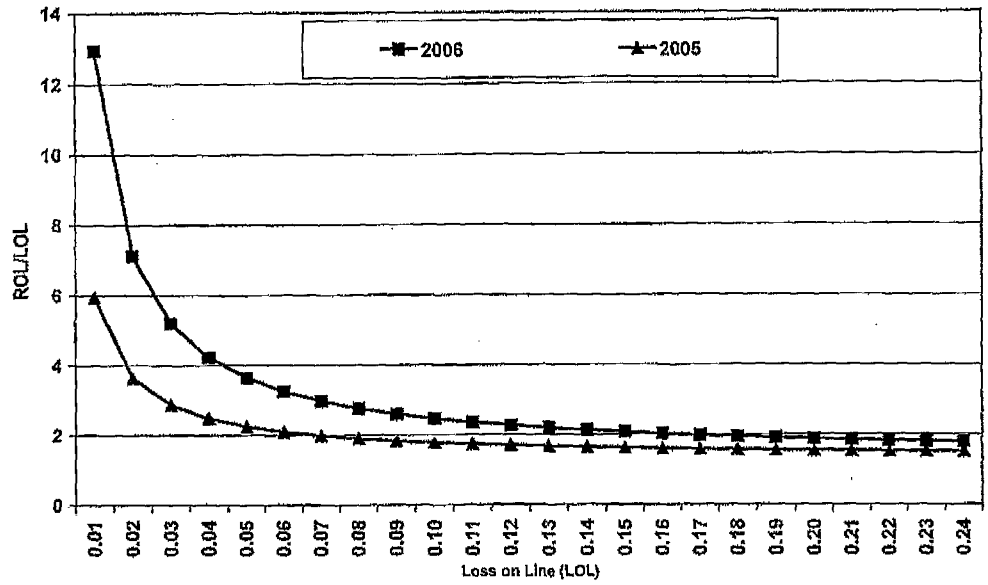
Source: Lane Financial LLC.

Comparison of CAT bond and catastrophe reinsurance pricing is difficult because of the general lack of systematic data on reinsurance prices. However, based on some unpublished data from Guy Carpenter, it is possible to provide a general indication of the comparative prices of CAT bonds and reinsurance. Guy Carpenter provided data on the relationship between the rate on line and the loss on line for catastrophe reinsurance. The rate on line (ROL) is defined as the reinsurance premium divided by the policy limit, and the loss on line (LOL) is the expected loss on the contract divided by the policy limit. The ratio of the ROL to the LOL is somewhat analogous to the ratio of the yield to expected loss on CAT bonds shown in Figure 9. The Guy Carpenter ROL and LOL data are based on average figures for Guy Carpenter clients buying reinsurance in 2005 and 2006 and are given separately for national primary insurers and regional primary insurers.

Like the CAT bond yield to expected loss ratios, the ratios of rates on line to expected loss on line for Guy Carpenter clients are significantly higher in 2006 than in 2005, reflecting the effects of Hurricanes Katrina, Rita, and Wilma. In addition, the ROL-to-LOL ratios are significantly larger for national insurers than for regional insurers. Finally, the ratios are lower for contracts with higher expected losses on line, reflecting the fact that policies with low expected LOL are covering the more risky upper tails of the loss distribution.

The ratios of ROL-to-LOL for national insurers in 2005 and 2006 are shown in Figure 10. The figure focuses on national insurers because the issuers of CAT bonds tend to be

FIGURE 10
Catastrophe Reinsurance Ratio of Rate on Line to Loss on Line, National Companies



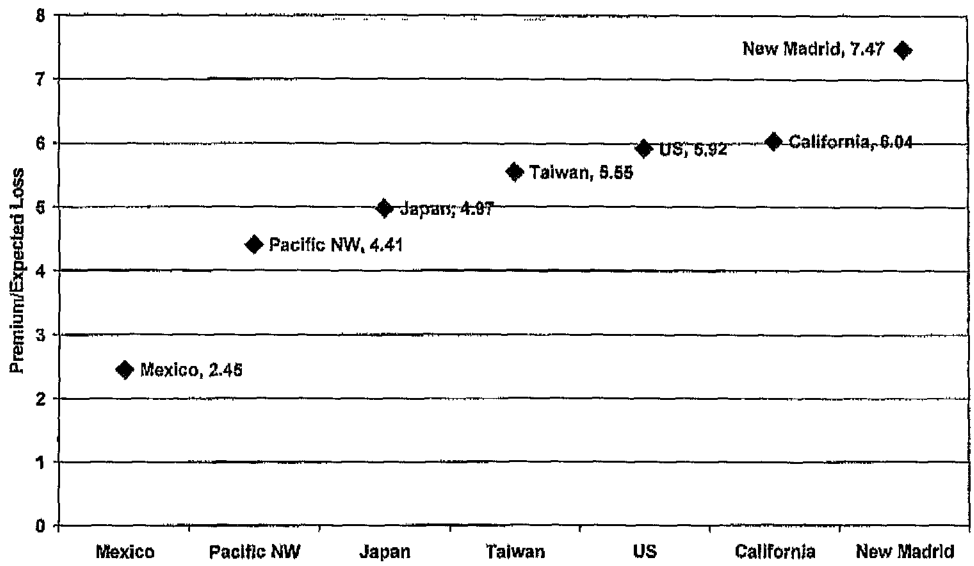
Source: Guy Carpenter.

large national and international firms. Thus, the most relevant comparison of CAT bond premia is with reinsurance prices for national insurers.

As shown in Figure 9, CAT bonds on average tend to have expected losses of between 1 and 3 percent of principal, and thus are most comparable to catastrophe reinsurance contracts with relatively low LOLs. As shown in Figure 10, the ROL-to-LOL ratios for LOLs of 1 percent, 2 percent, and 3 percent were 12.9, 7.1, and 5.2, respectively for national companies in 2006, and 5.9, 3.6, and 2.9, respectively for national companies in 2005. These compare with bond premium to expected loss ratios of around 3.3 in 2006 and 2.7 in 2005, based on averages of the four quarterly numbers for these years from Figure 9. Hence, even with the more normal pricing of 2005, CAT bonds clearly are "in the ballpark" in terms of pricing for national companies and also seem attractive relative to reinsurance in 2006. Hence, CAT bonds do not appear to be expensive relative to catastrophe reinsurance. Moreover, investment banks have succeeded in reducing transactions costs and speeding the time to market as they have gained experience with insurance-linked securitizations, also making the bonds more attractive to insurers and reinsurers.

For regional companies, at the 1 percent, 2 percent, and 3 percent LOL levels, the ROL-to-LOL ratios were 2.9, 2.5, and 2.3, respectively in 2006, and 2.4, 2.0, and 1.9, respectively in 2005. Thus, CAT bond prices look less attractive relative to reinsurance for regional companies. However, because regional firms have not been active in the CAT bond market, it is not clear what the bond premia would be for these firms.

FIGURE 11
Spreads on Selected Earthquake Bonds



Source: Cardenas (2006).

Another comparative indication of trends in CAT bond spreads is provided by a comparison of the Mexican CAT bonds with previously issued earthquake bonds. This comparison is provided in Figure 11, which shows the spreads on the Mexican bonds along with spreads on a representative selection of prior earthquake bonds. It is clear that the spreads on the Mexican bonds are very low in comparison to the prior bonds. This illustrates two phenomena, which cannot be precisely separated in terms of their influence on the spreads: (1) The Mexican bonds are more recent than the other bonds shown in the table, and, as indicated above, CAT bond spreads have been declining. (2) The Mexican bonds are valuable to CAT bond investors for diversification purposes because they cover a previously unsecuritized area of the world and permit investors to diversify their current large proportionate exposure to U.S. hurricane risk.

It is also relevant to compare CAT bond yields relative to yields on comparably rated corporate bonds. This comparison has been performed in MMC Securities (2007). The results show that BB CAT bond yields were comparable to yields on BB corporate bond yields from 2001 up until the time of Hurricane Katrina in 2005. Yields on CAT bonds exceeded yields on BB corporates during most of the period from the September of 2005 through February of 2007, although the gap had narrowed considerably by the end of the period. At the peak, yields on CAT bonds were 2 to 3 percent higher than the yields on BB corporates. Nevertheless, considering the magnitude of reinsurance prices in 2006 and the uncertainty created by Katrina and other recent catastrophes, the CAT bond market seems to have weathered the storms in very good shape.

REGULATORY, ACCOUNTING, TAX (RAT), AND RATING ISSUES

Prior discussions of alternative risk finance have indicated that regulatory and accounting treatment of CAT bonds and other risk financing solutions pose impediments to the growth of the market. However, industry experts interviewed by the author indicate that regulatory and accounting issues do not pose a material impediment to the growth of the market at the present time, and the statistics on market size and growth clearly seem to bear this out. Although a complete treatment of the regulatory, tax, and accounting issues are beyond the scope of this article, this section provides a few observations on the relevant issues, primarily to provide suggestions for future research that might be conducted on these topics.

Regulatory Issues

Some prior commentators have argued that CAT bonds have mostly been issued offshore for regulatory reasons and that the lack of onshore issuance represents a barrier to market developments. The argument is that encouraging onshore issuance might reduce transactions costs and facilitate market growth. However, industry experts interviewed by the author disagree with this point of view. They argue that the offshore jurisdictions, including Bermuda, the Cayman Islands, and Dublin, provide low issuance costs and high levels of expertise in the issuance of insurance-linked securities. Transactions costs for the onshore CAT bonds that have been issued generally have been higher than for offshore issues. Thus, whereas issuance of securities onshore (e.g., in the United States) probably would be a favorable development in the long run, the off-shore jurisdictions perform very effectively and efficiently in handling the issuance and settlement of insurance-linked securities.

Prior commentators have argued that nonindemnity CAT bonds currently face uncertain prospects with respect to regulatory treatment. The argument was that regulators are concerned about basis risk and the potential use of securitized risk instruments as speculative investments. As a result, it was argued that regulators may deny reinsurance accounting treatment for nonindemnity CAT bonds, impeding the development of the market. However, industry experts interviewed by the author indicate that regulatory treatment does not presently pose a significant obstacle to market development. Market participants have found a variety of structuring mechanisms to blunt regulatory concerns about alternative risk financing with respect to nonindemnity CAT bonds. For example, contracts can be structured to pay off on narrowly defined geographical indices or combinations of indices that are highly correlated with the insurer's losses. Concerns about speculative investing can be addressed through dual-trigger contracts that pay off on an index but where the insurer cannot collect more than its *ultimate net loss*, a familiar reinsurance concept equal to the insurer's total loss from an event less collections under reinsurance contracts.[6]

Even though regulation does not seem to pose a significant barrier to the development of the market at the present time, it remains true that the United States generally takes a heavy-handed, intrusive, and inflexible approach to insurance regulation. U.S. insurance

[6] This dual-trigger approach was developed in the market for *industry loss warranties*, which is a segment of the reinsurance market offering this type of contract (McDonnell, 2002).

regulation does not place sufficient reliance on the market as self-regulator of insurers and reinsurers. Regulation should primarily be designed to ensure transparency of insurance and reinsurance transactions, relying on the market to enforce appropriate behavior by insurers. Instead, U.S. regulation takes a "we must approve or disapprove everything" approach. It would be helpful to the efficiency of insurance markets in general if regulators were to adopt a more flexible regulatory approach. In the area of risk-linked securities, it would be helpful if regulators were to codify the rules and regulations relating to the statutory accounting treatment of various types of risk-linked securities and avoid imposing any unnecessary regulatory impediments in the future.[7]

Tax Issues

According to industry experts, offshore CAT bonds do not create taxation problems for sponsors. There generally are no income, corporate, withholding, or other significant taxes in offshore jurisdictions that apply to CAT bonds. The bond's SPRs are also not taxable for U.S. federal income tax purposes, provided that they are not held to be "engaged in a U.S. trade or business." Although no systematic information is available, anecdotal information suggests that so far offshore CAT bond SPRs have not been held to be engaged in a U.S. trade or business.

Consequently, the main tax issue involving CAT bonds for U.S. investors is the treatment of the bond premia under U.S. income tax law. The tax status is currently somewhat ambiguous given that neither the Tax Code nor the Internal Revenue Service addresses the tax treatment of income received from CAT bonds. Reportedly, bonds are presently being treated by many U.S. taxpayers who invest in the bonds as passive foreign investment companies (PFICs). Accordingly, income from CAT bonds is currently being included in taxable income as dividends rather than interest. U.S. sponsors also reportedly have been deducting premium payments on offshore bonds for income tax purposes, i.e., bond interest is currently treated in the same way as reinsurance premiums.

Dissemination of Information on Bonds

Although the ultimate objective should be the development of a public market for CAT bonds, privately placed bonds are likely to continue to play an important role. However, current securities regulations discourage the dissemination of information about private placements. Hence, market development is being impeded to the extent that information on existing bonds is not generally available. This discourages research by potential bond sponsors and by third parties such as academicians who might add significant value to the discussion.

Under current securities regulations, bond prospectuses for privately placed bonds can be distributed only to investors falling under the definition of accredited investors (or qualified investors) under Securities and Exchange Commission Regulation D. This class consists mainly of institutional investors and high net worth individuals. The rules are designed to prevent the sale of securities to the general public that have not gone through the Securities and Exchange Commission (SEC) registration process for public securities

[7] The NAIC has a model law on SPVs. For discussion of some of the issues, see Grace, Klein, and Phillips (2001).

issuance. However, the rules also have the unintended consequence of inhibiting research on CAT bonds.

The SEC rules should be changed to allow sponsors to distribute bond prospectuses to researchers who are not necessarily accredited investors. This could be done by posting the prospectuses on a repository maintained by an appropriate governmental entity. The repository could clearly indicate that the posting of the prospectuses on the site does not constitute an offer to sell and could require researchers downloading documents from the site to sign a strict users agreement. It still would not be possible for sponsors to sell bonds to the general public without appropriate registration with the SEC, but this change would make the bond prospectuses available for researchers.

Issues to Be Explored

What actions could be taken that would facilitate the further expansion of the market for insurance-linked securities? Although most of the regulatory issues mentioned in earlier discussions of CAT bonds are not problematical at the present time and the insurance-linked securities market is growing rapidly, there are some issues/reforms that may be able to enhance market development in the future. Several issues are mentioned here in the spirit of providing suggestions for further research on market development and efficiency.

Insurance regulators in key jurisdictions such as the United States, the European Union, and Japan could mandate catastrophe loss reporting for events above a given industry threshold such as $1 billion. Reporting could be done to a government agency or to a private organization such as Property Claims Services (PCS) in the United States. The mandate should require that data be reported in a significant amount of detail, probably more detail than presently provided by PCS. This would solve an important current problem, i.e., the lack of a PCS-equivalent index for the European Union and Japan, and would enhance the market by providing more information on U.S. losses. Regulators could work with knowledgeable insurers and reinsurers to design the data reporting specifications and address technology issues.

Until a loss turns into a recoverable, the quality of the reinsurance counterparty is effectively ignored by regulators. This is the case, for example, in the U.S. risk-based capital system, where the charges for reinsurance are not graded by reinsurer credit quality. Explicitly incorporating reinsurance credit quality into regulatory capital calculations and related regulatory credit evaluations has the potential to provide an important boost to the insurance-linked securities market as well as improving insurance solvency regulation in more general terms.

As pointed out in Cummins (2007), personal lines insurers in the United States face price regulation in several key catastrophe-prone states. In general, insurers reevaluate loss distributions and file for price increases following major catastrophes, reflecting changing estimates of expected losses and higher reinsurance premiums. Unfortunately, the political reality is that regulators are most reluctant to allow price increases at precisely the times when insurer loss expectations and reinsurance prices are increasing most rapidly. The best solution to this problem would be to deregulate prices at the state level so that primary insurers would not be caught in this price-cost bind. Short of deregulating prices, regulators could help ease the problem by giving primary insurers credit for locking in multi-year pricing and capacity by issuing insurance-linked securities, in view of the fact that a vast majority of reinsurance policies renew annually whereas insurance-linked securities often cover multi-year terms.

Some industry experts observe that the application of the Employee Retirement Income Security Act (ERISA) to CAT bond collateral trusts may make it marginally more difficult to attract foreign investors. As with many ERISA issues, the matter is complex and would benefit from some thorough research into the current rules and any law changes or U.S. Department of Labor rule changes that might clarify the situation and point the way to potential reforms.

CONCLUSIONS

The CAT bond market is thriving and seems to have reached "critical mass." The market achieved record bond issuance in 2005, 2006, and 2007. Bond premia have declined significantly since 2001 and the bonds now are priced competitively with catastrophe reinsurance. Even following Hurricane Katrina, bond premia were roughly comparable to yields on similarly rated corporate bonds. The amount of risk capital raised through CAT bonds has been growing, and the bonds now account for a significant share of the property-catastrophe reinsurance market. The bonds have an especially important role to play for high coverage layers and in the retrocession market. Considering CAT bonds, swaps, and industry-loss warranties, many experts believe that these alternative risk transfer devices now account for more than half of the property insurance retrocession market and are of growing importance in other parts of the market.

Regulatory and accounting issues such as the regulatory accounting treatment of non-indemnity CAT bonds and the issuance of most bonds offshore, which have been cited as impediments to the development of the market, do not presently seem to pose serious problems. However, there are a number of issues/reforms that should be explored to provide ways in which public and private institutions can facilitate market development. These include fostering better reporting of catastrophe losses to facilitate the development of better CAT loss index products. Solvency regulation should be adapted to recognize the credit quality of reinsurance receivables and give recognition to the full collateralization provided by CAT bonds. Primary insurance prices should be deregulated in the United States, and primary insurers should receive credit from regulators for entering into contracts that provide multi-year pricing and capacity through either insurance-linked securities or conventional reinsurance. Other issues to be investigated include the applicability of ERISA to CAT bond collateral trusts and the U.S. GAAP and statutory accounting treatment of triggers employed in industry loss warranties and similar contracts. Finally, issuers of CAT bonds should be required to make available bond prospectuses to researchers who could provide valuable analysis of catastrophe risk financing. The prospectuses could be made available on a government web site and users would be required to sign a users agreement proving penalties for misuse of the information contained in the documents.

The future looks bright for the insurance-linked securities market. CAT bonds, swaps, sidecars, industry loss warranties, and other innovative products will play an increasingly important role in providing risk financing for large catastrophic events. Event-linked bonds are also being used increasingly by primary insurers for lower layers of coverage and noncatastrophe coverages such as automobile and commercial liability insurance. It remains to be seen whether CAT futures and options will play an important role in catastrophe risk management in the years to come. Basis risk and counterparty

credit risk, as well as the need to educate insurance industry participants, are the primary impediments to the success of these contracts.

REFERENCES

A.M. Best Company, 2006, 2006 *Annual Global Reinsurance Report: Reinsurers Humbled, but Most Not Broken, by Hurricane Losses* (Oldwick, NJ).

Cardenas, V., 2006, Sovereign Financial Disaster Risk Management: The Case of Mexico, PowerPoint presentation, Ministry of Finance and Public Credit, Mexico City, Mexico.

Cardenas, V., S. Hochrainer, R. Mechler, G. Pflug, and J. Linnerooth-Bayer, Forthcoming, Sovereign Financial Disaster Risk Management: The Case of Mexico, *Environmental Hazards*.

Cowley, A., and J. D. Cummins, 2005, Securitization of Life Insurance Assets and Liabilities, *Journal of Risk and Insurance*, 72: 193-226.

Cummins, J. D., 2005, Convergence in Wholesale Financial Services: Reinsurance and Investment Banking, *The Geneva Papers* 30 (April): 187-22.

Cummins, J. D., 2006, Should the Government Provide Insurance for Catastrophes, *Federal Reserve Bank of St. Louis Review*, 88: 337-379.

Cummins, J. D., 2007, Reinsurance for Natural and Man-Made Catastrophes in the United States: Current State of the Market and Regulatory Reforms, *Risk Management and Insurance Review*, 10: 179-220.

Cummins, J. D., D. Lalonde, and R. D. Phillips, 2004, The Basis Risk of Index-Linked Catastrophic Loss Securities, 2004, *Journal of Financial Economics*, 71: 77-111.

Froot, K. A., 2001, The Market for Catastrophe Risk: A Clinical Examination, *Journal of Financial Economics*, 60: 529-571.

Grace, M. F., R. W. Klein, and R. D. Phillips, 2001, Regulating Onshore Special Purpose Vehicles, *Journal of Insurance Regulation*, 19: 551-590.

Guy Carpenter, 2005a, *The Growing Appetite for Catastrophe Risk: The Catastrophe Bond Market at Year-End 2004* (New York).

Guy Carpenter, 2005b, *The World Catastrophe Reinsurance Market* (New York).

Guy Carpenter, 2006a, *The Catastrophe Bond Market at Year-End 2005: Ripple Effects from Record Storms* (New York).

Guy Carpenter, 2006b, *U.S. Reinsurance Renewals at January 1, 2006* (New York).

Guy Carpenter, 2007a, *The World Catastrophe Reinsurance Market 2006: Steep Peaks Overshadow Plateaus* (New York).

Guy Carpenter, 2007b, *U.S. Reinsurance Renewals at January 1, 2007: Smooth Sailing* Ahead? (New York).

Harrington, S. E., and G. Niehaus, 2003, Capital, Corporate Income Taxes, and Catastrophe Insurance, *Journal of Financial Intermediation*, 12: 365-389.

Lane, M., 2007, *Of Sidecars and Such* (Wilmette, IL: Lane Financial).

Lane, M., and R. Beckwith, 2005, *The 2005 Review of the Insurance Securitization Market: GAME ON!* (Wilmette, IL: Lane Financial).

Lane, M., and R. Beckwith, 2006, *How High Is Up: The 2006 Review of the Insurance Securitization Market* (Wilmette, IL: Lane Financial).

Lane, M., and R. Beckwith, 2007a, *That Was the Year that Was! The 2007 Review of the Insurance Securitization Market* (Wilmette, IL: Lane Financial).

Lane, M., and R. Beckwith, 2007b, *Developing LFC Return Indices for Insurance Securitizations* (Wilmette, IL: Lane Financial).

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, 76–86.

McDonnell, E., 2002, IndustryLoss Warranties, in: M. Lane, ed., *Alternative Risk Strategies* (London: Risk Books).

MMC Securities, 2007, *The Catastrophe Bond Market at Year-End 2006: Ripples Into Waves* (New York).

Mocklow, D., J. DeCaro, and M. McKenna, 2002, *Catastrophe Bonds*, in Morton.

Swiss Re, 2001, *Capital Market Innovation in the Insurance Industry Sigma No. 3/2001* (Zurich, Switzerland).

Swiss Re, 2006, *Securitization: New Opportunities for Insurers and Investors* (Sigma No. 7/2006 Zurich, Switzerland).

Swiss Re, 2007a, *Natural Catastrophes and Man-Made Disasters in 2006: Low Insured Losses* (Sigma No. 2/2007 Zurich, Switzerland).

Swiss Re, 2007b, *Insurance Linked Securities Market Update* (August).

Takeda, Y., 2002, Risk Swaps, in: M. Lane, ed., *Alternative Risk Strategies* (London: Risk Books).

CAS Exam 9 Study Kit

"Asset/Liability Management Strategies for Property and Casualty Companies"

by P.D. Noris

MORGAN STANLEY

Asset/Liability Management Strategies for Property & Casualty Companies

Peter D. Noris, C.F.A.
Product Manager

May 1985

TABLE OF CONTENTS

INTRODUCTION

In addition to being an "insurance" risk intermediary, a Property and Casualty (P&C) insurance company has become a "financial" risk intermediary. Gone are the days when an insurer could be paid *solely* for the function of reducing the disutility of the risk of each insured's loss. The high rates of return that may be earned on insurance premiums while the insurer is waiting for losses to occur, and then be paid, have caused insurers to be dragged, perhaps unknowingly, into the role of financial intermediary. The additional investment risks inherent in this role have long been known to other financial intermediaries, such as commercial banks (and more recently to life insurance companies[1]), but P&C companies are relatively new to this game. It is true that the P&C business is a unique bundle of risks, but many of the techniques developed for other financial intermediaries can be adapted to help P&C insurers manage these varied risks more profitably.

The purpose of this paper is to illustrate a typical P&C insurance company and to suggest how some asset/liability management techniques may be used in order that company management may better control its own destiny. Section I of this paper develops a computer model of the entire P&C insurance industry that is used to illustrate the dynamics of insurance underwriting and investment profitability. The nature of the industry's liabilities, the time profile of their insurance claims, is then explored in Section II. Once the nature of the assets and liabilities is understood, a methodology is developed in Section III which allows us to place a market value on any P&C insurance company. The ability to manage the market value of an insurance company is achieved in Section IV by computing the firm's sensitivity to changes in interest rates. By raising or lowering the firm's sensitivity to interest rates, company management can partake in structured risk taking, choosing when to emphasize investment risk or when to emphasize insurance underwriting risk.

Sections V and VI show the process that would be involved in developing and implementing an investment strategy for a P&C insurer. An investment strategy is illustrated by hypothesizing a number of possible goals of company management and then developing a simple strategy to achieve those goals. The computer model developed previously is then used, in Section VII, to test the historical effectiveness of the strategy, which is shown to be very effective.

1 See Toevs and Haney (1984), and Tilley and Jacob (1984).

SECTION I

THE TYPICAL P&C INSURANCE COMPANY: AN ILLUSTRATION

This paper develops a model insurance company. The word "model" does not imply that it is an ideal company, rather it refers to the fact that the company is a computer model created using industry-wide data from "Best's Aggregates & Averages." This methodology could certainly be applied to any individual company (and this will be a major thrust of Morgan Stanley's analytical effort in the P&C industry), but the aggregates are meant to be relevant to most companies in the industry today. While the illustrations are developed from industry data, some liberties have been taken for the sake of clarity.

XYZ Casualty, as we will refer to the model company, is constructed to be a primary writer of five lines of business. Sales have been indexed to equal the industry's total sales for each line (percentage of total 1972 premiums shown in parentheses): Auto Liability (39%), Auto Physical Damage (22%), Multiple Peril (17%), Workers' Compensation (13%) and General Liability, including Medical Malpractice (9%). For purposes of our study the company came into "existence" in 1972 and has written business through year-end 1983. By choosing the five largest lines of P&C insurance sales (accounting for over 70% of total industry premiums) XYZ has experienced the same trends in premium growth and underwriting profitability ratios as the aggregate P&C industry, as shown in Figures 1 and 1A.

XYZ Investment Policy

XYZ had an investment policy that changed over time but was relatively stable from any one year to the next. In 1972 the first step in implementing this policy was to determine the amount of the Company's Statutory Policyholder's Surplus.[2] The amount of surplus dictated the amount of the firm's holdings in common and preferred equities. While XYZ desired a high percentage of its assets to be invested in stocks (possibly its management felt that equities would achieve a long-term real rate of return), the company could not risk an accounting reduction in its surplus account that would restrict the amount of premiums that could be written (insurance regulators typically limit a company's premium writings to no more than 3 or 4 times surplus). Because common stocks and non-sinking fund preferreds are the only assets that XYZ is required to mark-to-market at the end of each accounting period, it was felt that the insurer should somewhat limit its policyholders' (and shareholders', if it is a stock company) risk of any sharp decline in stock prices. Investment policy in 1972 was to invest no more than 100% of policyholder's surplus in equities, an amount thought to be conservative at the time. We will see that this policy changed radically later in the decade. The annual return on the equity portion of XYZ's investment portfolio is represented by the annual total-return[3] on the S&P 500 Index.

2. Many definitions are contained in the Appendix.

3. Includes both capital appreciation and dividend income.

Figure 1

Net Premiums Written

(Indexed 1972 = 100)

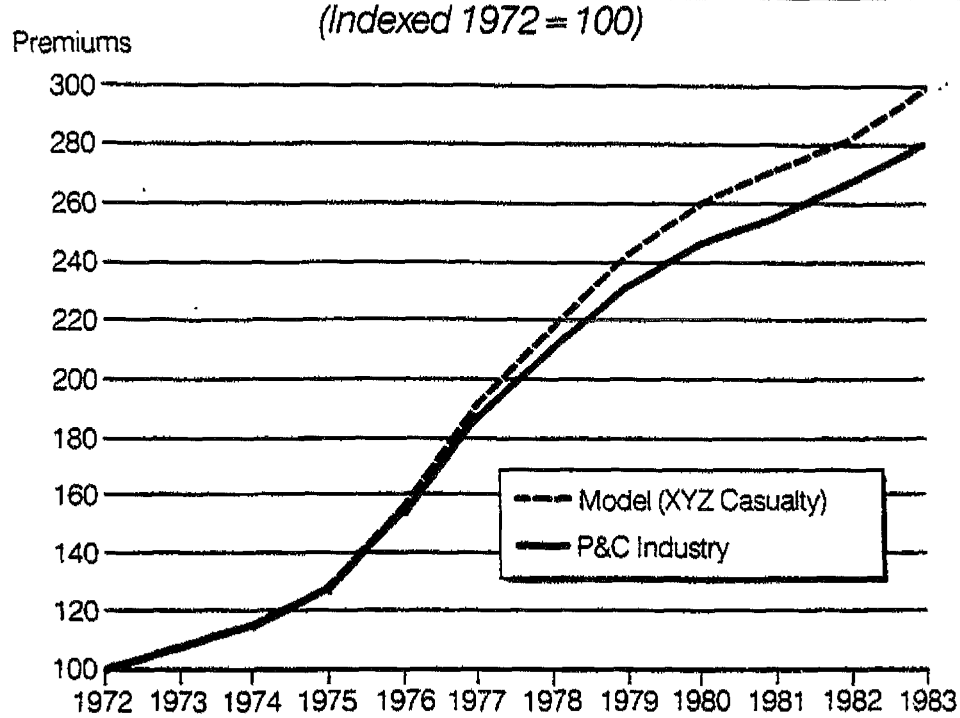
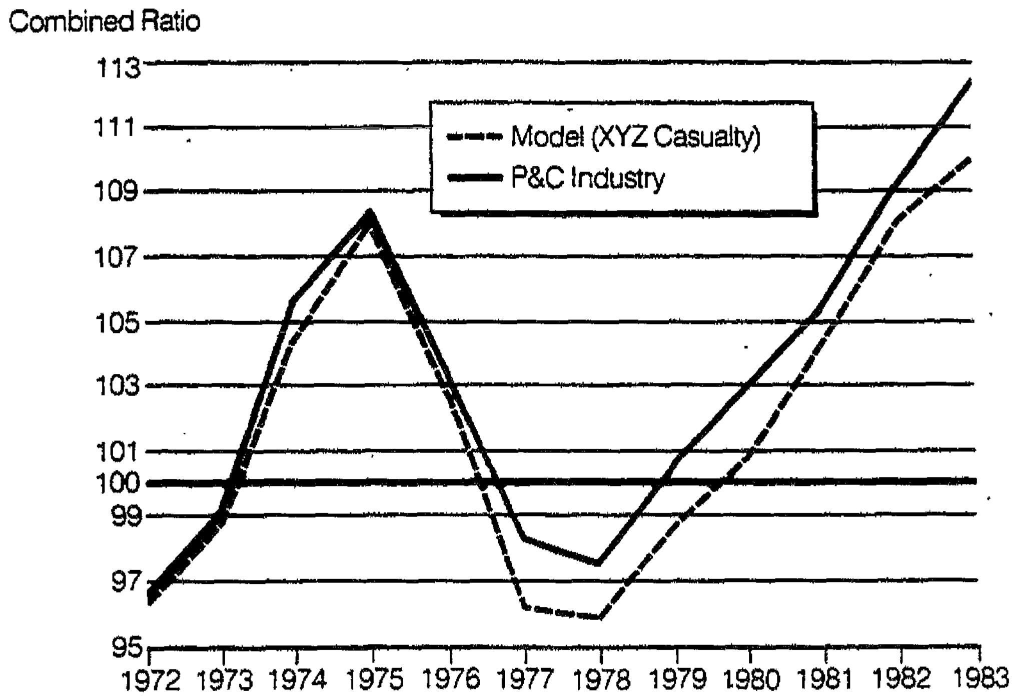


Figure 1A

Combined Ratio



The balance of the investment portfolio was placed in instruments that could be held at amortized cost. These instruments included taxable and tax-exempt bonds and taxable money-market instruments. Because of the high effective tax rate of the industry, caused by the profitability of insurance underwriting, and the expectation that this high rate would continue, XYZ decided to place the majority of its remaining assets in tax-exempt Municipal bonds, thus sheltering any investment earnings from income tax. The final portion of assets was invested in taxable bonds, divided between long-term bonds and money-markets, the latter serving as a liquidity reserve.

The decision as to sector and maturity of the fixed-income holdings came not so much from the perspective of what types of insurance liabilities the company was writing, but more from "market" decisions based upon the Investment Committee's analysis of what sectors looked cheap and what part of the yield curve offered value. XYZ felt that so long as the company was looked at from a "going concern" perspective, asset/liability matching need not play a major role in the investment allocation decision. While it was true that most policies' premiums would be available for investment for just a few years, XYZ would be renewing these insurance policies (or ones similar to them) and this would create a stable core of funds that could be invested to any desired part of the maturity spectrum, no matter how long or short. Day-to-day cash needs and any catastrophic losses could be handled by the approximately 5% of assets that were held in money-market instruments. So went the conventional wisdom of the day.

This rationale of largely ignoring liability considerations led to the practice of purchasing predominately long-term (at least 20 years to maturity) bonds, particularly in the Municipal sector where there has traditionally been a very steep positively-sloped yield curve. In our model, which allows for investment at only year-ends, XYZ's practice was to buy 20-year Municipal bonds with yields equal to Moody's Bond-Buyer Index, 20-year taxable bonds with yields equal to Moody's Corporate Bond Composite Index and one-year money-market instruments with yields equal to the annual return on one-year Treasury-Bills. The percentage of assets invested in each asset, as well as the various year-end yields (for bonds) or yearly book rates of return (for stocks and money-markets), are shown in Table I.

As can be seen in Table I, the only major shift that XYZ made in investment policy occurred in 1974 when the percentage of assets devoted to equities was dramatically reduced (from 32% of assets to 19%). One inadvertent reason for this reduction was the disastrous stock market performance in 1974, the S&P fell over 25%, which translated into a reduced equity, and surplus, position. But XYZ also consciously decided to reduce its exposure to equities in addition to that deterioration; 1973 had also been a poor year for equities, declining 15%, and 1974 was shaping up to be a poor year on the underwriting side, with the company's combined ratio coming in at over 104%. Because prospects for 1975 underwriting results did not look any rosier, XYZ decided to reduce its exposure to, and limit the balance sheet impact of, any potential further deterioration in the stock markets. While this timing was unfortunate because the S&P actually rose 25% in 1975, the reduction in risk was necessary from the standpoint of prudency. The reduction in equities as a percent of policyholder's surplus to the 60% range remained in effect through 1983.

Table I

XYZ Casualty

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Balance Sheet (percent)												
Assets												
Money Market	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%
Taxable	10	23	36	38	38	41	40	38	34	36	35	34
Tax-Exempt	40	40	40	40	40	40	40	40	40	40	40	40
Common & Pfd	45	32	19	17	17	14	15	17	21	19	20	21
Total Assets	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Liabilities + Surplus												
Loss Reserves	50%	64%	76%	76%	76%	76%	74%	72%	68%	69%	67%	65%
Surplus	50	36	24	24	24	24	26	28	32	31	33	35
Total Liab. + Surplus	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Stock to Surplus Ratio	91%	91%	78%	72%	68%	58%	57%	61%	65%	61%	59%	60%
Annual Rate of Return												
Common & Pfd: Div.	N/A	3.46%	5.25%	4.08%	3.77%	4.91%	5.28%	5.23%	4.54%	5.41%	4.88%	4.30%
Cap.	N/A	-18.32	-30.50	30.07	18.30	-12.11	1.20	13.25	28.84	-10.99	15.91	17.87
Money Market	N/A	5.80	7.40	7.30	6.30	4.90	7.10	12.90	12.20	14.20	14.40	8.80
Year-End Yields												
Taxable Bonds	7.47%	8.05%	9.63%	9.57%	8.47%	8.54%	9.49%	11.35%	14.04%	15.38%	13.02%	13.01%
Tax-Exempt Bonds	5.11	5.16	7.08	7.29	5.83	5.66	6.61	7.23	9.76	13.30	9.56	9.76

Results of XYZ's Investment and Underwriting Practice

Table II shows yearly Income Statements and Balance Sheets for our model insurance company for the eleven-year period between 12/31/72 and 12/31/83. Other relevant statistics. such as average investment yield, stock to surplus ratio and bond portfolio book-to-market ratio are also shown in this table. The tax rate used in calculating the annual tax expense is the effective corporate tax rate for any taxable income (48% tax rate through 1979 and 46% thereafter).

One can see from Table II the extremely poor underwriting results of 1974-1975 and the recent years of 1982 and 1983. In these years our model company actually generated tax benefits from having underwriting losses which exceeded taxable investment income. These benefits helped the company to show positive net operating income in 1974, 1982, and 1983, and to almost break even in 1975. It is likely that our model would have shown positive net operating income in 1975 also, had our analysis gone back far enough to include tax benefits from underwriting losses in the late 1960s.

Likewise, Table II also shows the stellar years of 1977, 1978 and 1979, when strong underwriting gains combined with growing investment income to produce annual book returns-on-surplus[4] of 20% or more. But in some ways it was these halcyon days of the late 1970s that were hiding the sores that would begin to fester in the 1980s. As we will see in Section III, the accounting reports that were showing record profitability belied the chronic asset/liability mismatch of the industry.

A relevant exercise at this point is to see if the results of our model company behave like a typical P&C insurance company. Figures 2, 3 and 4 show how the annual performance of the model company, XYZ Casualty, compares to industry aggregates from Best's. Figure 2 shows the percent change in the model's and the industry's Statutory Surplus each year. Figure 3 shows annual return-on-surplus results. Figure 4 develops investment income as a percent of mean assets. It is readily apparent that our model insurance company looks very much like the insurance industry aggregates. Even though the model company exhibits slightly more volatile performance figures, which would be expected since we began with a clean balance sheet on 12/31/72 and are writing only five lines of business, the model results are generally in the right direction and of the same magnitude as the industry aggregates. The model's slight underperformance, as measured by return-on-surplus, can be explained by its higher effective tax rate compared to the industry (our model company does not have any tax loss carry-forwards from pre-1972 underwriting periods). The "fresh start" we gave our company would also explain the model's slightly higher average investment yields than those of the industry, which include investments from business that was written in much lower yield environments of years before 1972.

It is interesting to see that an industry as large and complicated as the Property and Casualty Industry can be very nearly replicated by a fairly simplified model that assumes only five lines of insurance, four available investments and allows inflow or outflow of funds only at year end. This is a fortunate circumstance, since it significantly simplifies Section IV of our analysis which looks at the impact of changing the firm's investment policy to account for the nature of the industry's liabilities.

4. ROS = Net Operating Income ÷ Beginning Surplus.

Figure 2

Change in Surplus

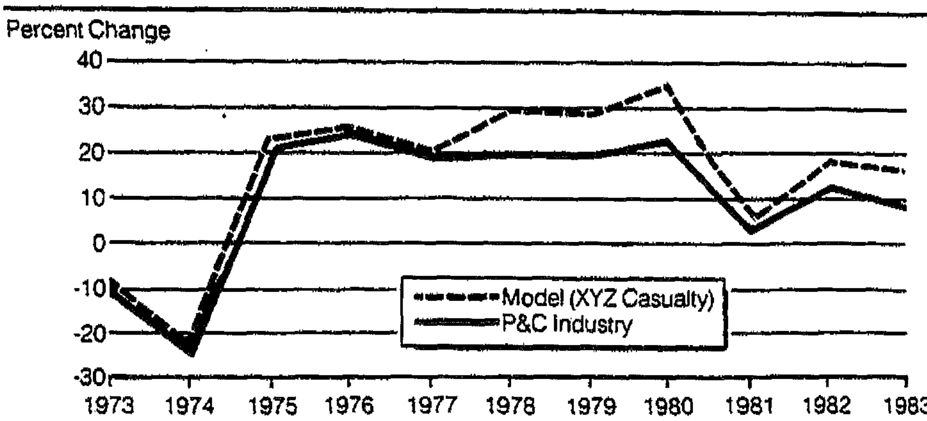


Figure 3

Return on Surplus

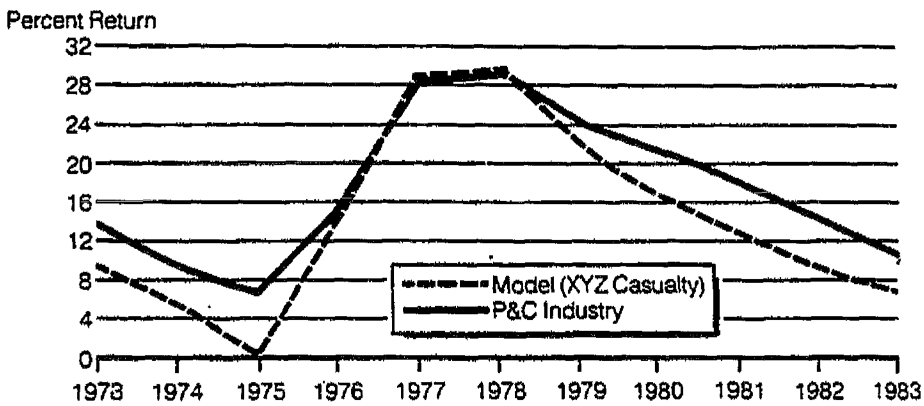


Figure 4

Investment Income as % of Mean Assets

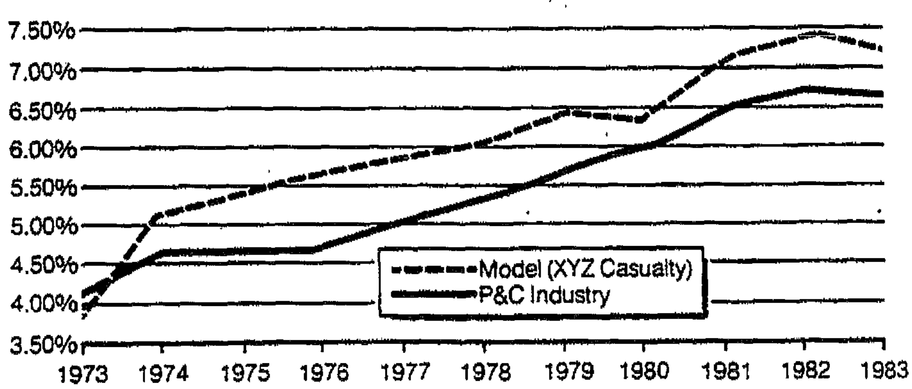


Table II

XYZ Casualty

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Income Statement												
Net Premium Earned												
Auto Liab	$11,383	$11,744	$11,889	$12,807	$15,122	$18,159	$20,073	$21,682	$22,934	$24,011	$25,610	$27,528
Auto Phys	6,408	6,864	7,106	7,399	8,962	10,961	12,484	13,969	15,466	16,469	17,579	19,178
Workers Comp	4,010	4,676	5,332	5,965	7,274	9,072	10,771	12,653	13,895	14,407	13,918	14,097
Multi Peril	5,083	5,941	6,815	7,751	9,382	11,309	13,162	15,082	16,544	17,562	18,828	20,014
General Liab	2,564	2,734	2,921	3,644	5,038	6,571	7,380	7,726	7,797	7,370	7,077	7,240
Total Premium	$29,448	$31,959	$34,063	$37,566	$45,724	$56,072	$63,870	$71,112	$76,636	$79,819	$83,012	$88,057
Expenses												
Losses Incurred	20,519	22,974	26,201	30,712	35,610	40,415	45,489	52,542	57,794	62,401	67,338	72,846
Underwriting Exp	7,842	8,613	9,339	9,862	11,355	13,546	15,763	17,696	19,430	20,915	22,412	24,069
Net Underwriting Income	$ 1,087	$ 372	($ 1,477)	($ 3,008)	($ 1,240)	$ 2,111	$ 2,618	$ 874	($ 588)	($ 3,498)	($ 6,738)	($ 8,858)
Investment Income												
Money Mkt	0	66	111	126	136	131	227	494	556	769	847	580
Taxable	0	164	529	1,084	1,462	1,807	2,289	2,724	3,132	3,500	4,309	4,902
Tax-Exempt	0	469	612	741	996	1,233	1,475	1,807	2,226	2,896	3,400	3,938
Dividend	0	360	510	263	275	433	478	595	715	1,221	1,099	1,105
Investment Exp	0	(76)	(99)	(114)	(143)	(176)	(211)	(253)	(301)	(357)	(388)	(435)
Pre-tax Investment Income	$ 0	$ 984	$ 1,663	$ 2,100	$ 2,728	$ 3,428	$ 4,257	$ 5,367	$ 6,328	$ 8,029	$ 9,266	$10,091
Pre-tax Operating Income	$ 1,087	$ 1,356	$ 186	($ 908)	$ 1,487	$ 5,539	$ 6,875	$ 6,242	$ 5,739	$ 4,532	$ 2,528	$ 1,233
Income Taxes	522	279	(413)	(899)	123	1,890	2,397	1,886	1,337	275	(831)	(1,677)
Net Operating Income	$ 565	$ 1,077	$ 599	($ 9)	$1,364	$3,649	$4,478	$4,356	$4,403	$4,257	$3,358	$2,910

Table II

XYZ Casualty

(continued)

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Balance Sheet												
Assets												
Money Market	$ 1,146	$ 1,495	$ 1,722	$ 2,159	$ 2,667	$ 3,200	$ 3,829	$ 4,553	$ 5,412	$ 5,885	$ 6,589	$ 7,353
Taxable	2,197	6,731	12,492	16,448	20,521	26,155	30,748	34,341	36,962	42,218	46,778	50,350
Tax-Exempt	9,172	11,958	13,773	17,274	21,339	25,603	30,633	36,426	43,298	47,080	52,715	58,822
Common & Pfd	10,414	9,711	6,446	7,304	8,820	9,050	11,373	15,746	22,572	22,517	25,705	30,529
Total Assets	$22,929	$29,896	$34,433	$43,186	$53,348	$64,009	$76,583	$91,066	$108,244	$117,700	$131,787	$147,054
Liabilities & Surplus												
Loss Reserves	$11,465	$19,262	$26,162	$32,986	$40,447	$48,527	$56,515	$65,135	$73,370	$81,049	$88,196	$95,960
Surplus	11,465	10,634	8,271	10,200	12,901	15,482	20,068	25,931	34,874	36,651	43,592	51,095
Total Liab. & Surplus	$22,929	$29,896	$34,433	$43,186	$53,348	$64,009	$76,583	$91,066	$108,244	$117,700	$131,787	$147,054
Stock to Surplus Ratio	91%	91%	78%	72%	68%	58%	57%	61%	65%	61%	59%	60%
Inv. Inc. % of Assets	N/A	3.73%	5.17%	5.41%	5.65%	5.84%	6.06%	6.40%	6.35%	7.11%	7.43%	7.24%
Ratio of Market Value to Book:												
Taxable Bonds	100%	98%	92%	95%	104%	102%	95%	84%	71%	71%	86%	87%
Tax-Exempt Bonds	100%	100%	84%	87%	102%	104%	95%	92%	78%	63%	88%	88%

SECTION II

THE NATURE OF P&C LIABILITIES

P&C liabilities have a characteristic time profile—a premium that is received today will be paid out in the form of a claim some time in the future. The timing and size of this claim payment depends upon the nature of the line of business—automobile damage claims settle rather quickly and cannot exceed the value of the automobile while medical malpractice claims may take years to litigate and their ultimate size is somewhat at the whim of a jury. Any individual claim may settle sooner or later than expected (no one knows when a particular ship will sink), but reasonably stable payment patterns may be determined in the *aggregate*. This is one of the principles on which insurance practice is based.

Once the payment pattern for a line of business has been determined, various *immunization* techniques may be used to increase the probability that investment results are sufficient to achieve a profitable return for that line of business. As mentioned in Section I, traditional investment practice has largely ignored the nature of liability payment patterns, relying instead upon a hoped-for stability in investment markets so that a previously purchased long-term investment could be used to back the new year's written coverages (after the new year's premiums were, in effect, used to pay off claims on previous years' coverages). This practice also relied upon a stability in insurance pricing from one policy year to the next, in that the income from last year's investments will be adequate only if this year's pricing does not assume any extra income from higher market rates of interest. We are all too much aware of the recent instability of the investment and insurance markets. In such an environment, investment strategies that do not account for the nature of the liability side of the balance sheet are likely to cause great financial distress.

Typical Liability Payment Patterns

The five insurance lines of XYZ Casualty each have a characteristic payment pattern, known as their "loss development."[5] For purposes of this exposition, we have assumed that loss reserve data published for the P&C industry are a reasonable basis for projections and illustrations. These patterns were gleaned from two "Best's Insurance Management Reports" articles dealing with 1983 casualty loss reserves.[6] In general, loss payment patterns for any individual company are developed in Schedules O and P of a company's Statutory Annual Statement. In addition, company actuaries are routinely involved in developing their best estimates as to future loss payments resulting from each year's coverages.

Automobile Physical Damage

Coverage for Automobile Physical Damage is a classic example of what is known as a "short-tailed" line of business. The "tail" refers to the length of time it takes to settle or pay off a claim once it has occurred. Typical industry data shows that, 82% of the time, an insured automobile accident will be paid off within one year.[7] In fact, the tail on Auto Physical Damage

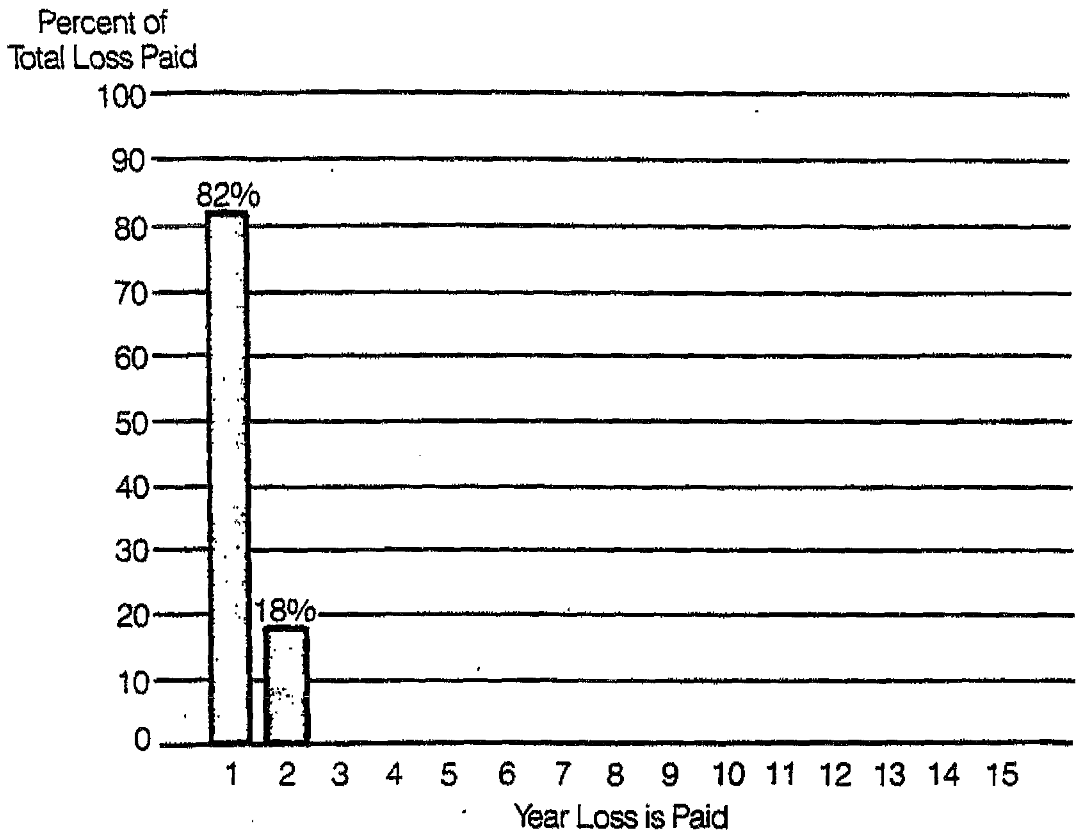
5. The amount by which actual losses vary from original projections is also sometimes referred to as "loss development."

6. "1983 Casualty Loss Reserves," *Best's Insurance Management Reports*, November 12, 1984 and December 24, 1984.

7. Loss patterns for Auto Physical Damage were not included in the Best's articles and have been developed from individual company Annual Statements.

Figure 5

Auto Physical Damage Loss Development



rarely extends beyond the second accident loss year and, for the sake of our analysis, we will assume that all losses are paid in years one and two.

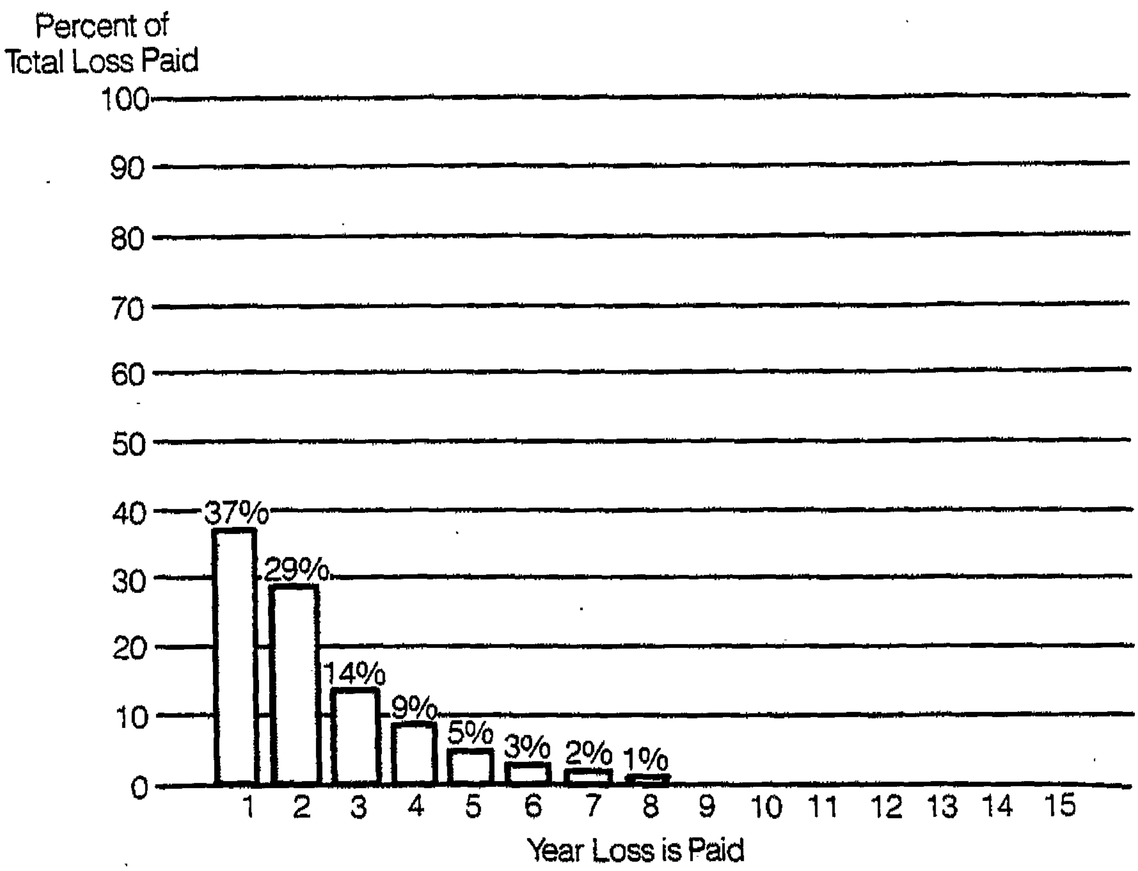
As a visual tool in comparing various lines of insurance business, we will graph the loss payment patterns for each line, placing the percent of total to be paid on the "Y-axis" and number of years since occurrence of the loss on the "X-axis." We can quickly see the short-tailed nature of Automobile Physical Damage (Figure 5).

Automobile Liability

Automobile Liability coverage has a longer tail than Physical Damage coverage. Liability claims are susceptible to negotiation and litigation and an insurance company may find that it has use of premium funds for many years until a claim eventually has to be paid. In many ways it was the industry's realization that long-tailed claims could result in substantial amounts of investment income that fostered the practice of "cash flow underwriting," which is explained in more detail in a later section. Best's data indicates that only 37% of Auto Liability claims will be paid within the first year and less than two-thirds of the claims by the end of the second year. Typical data from individual companies indicate that the tail extends about 8 years.

Figure 6

Auto Liability
Loss Development



An added risk of any long-tailed line is that eventual payments on claims may differ significantly from original projections. Many unknowns impact the ultimate size of a claim, not the least of which is the future rate of inflation. Actuaries must account for some amount of inflation of future claims in their loss development analyses, but the recent violent swings in inflation (particularly inflation of medical costs) have shown these assumptions to be suspect. There is also the additional impact of "social" inflation, as court awards rise higher and higher each day, that must somehow be accounted for. The topic of immunizing against an inflation rate is discussed later in this paper and, for the present, we will assume that the company actuary has accurately projected the future loss development pattern for each line of business.

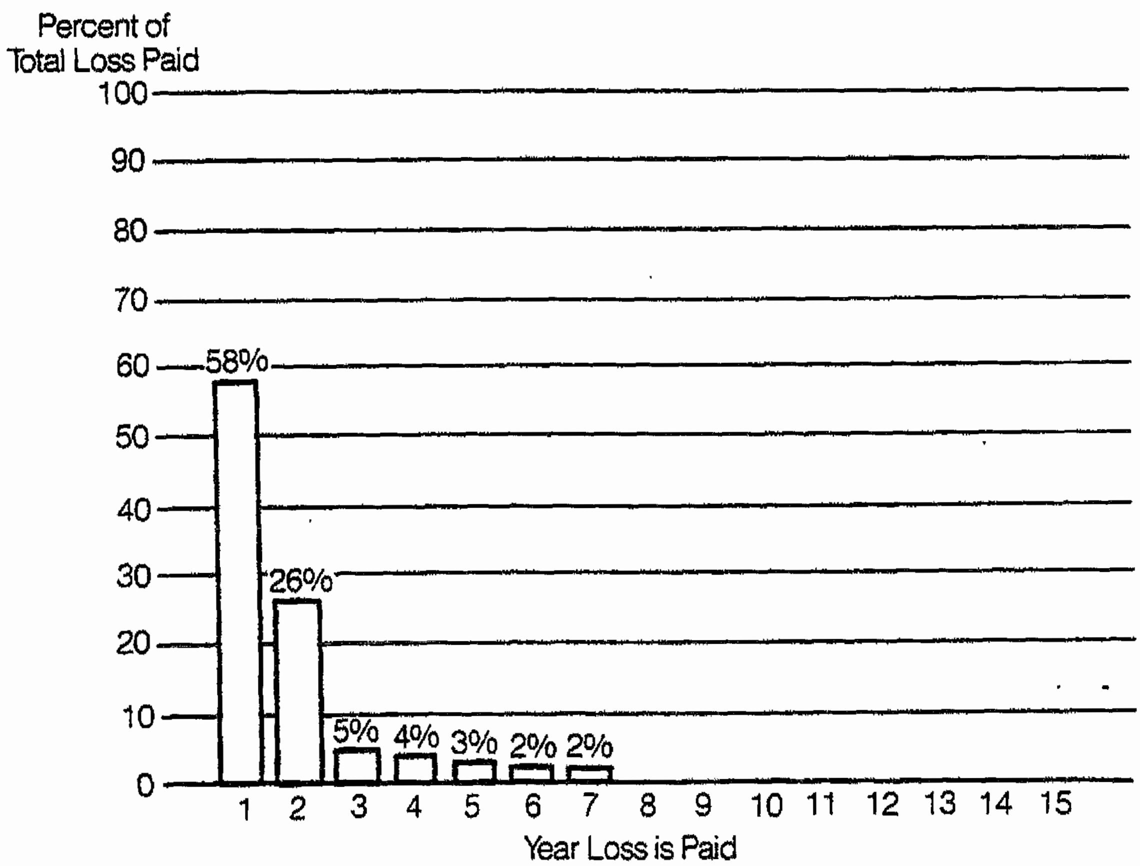
Multiple Peril

Multiple Peril insurance is a combination of both property (damage) and liability insurance. Perils that are covered include such things as fire, lightning, hail, windstorms, and explosions, and the insurance policy covers both the repair of physical damage as well as coverage for any liability claims that arise from the catastrophic events. The lines of business that are included in this category are farmowners, homeowners and commercial multiple peril in addition to ocean marine, aircraft, and boiler and machinery. Because Multiple Peril consists predominantly of property coverage, it has a shorter tail than pure liability coverage, with almost 96% of incurred losses being paid within 5 years. The shorter tail also means that there is little time for unanticipated occurences to impact the size of the eventual claims, thus the line exhibits relatively stable loss development statistics over the years as losses develop in accordance with the original projections.

Another statistic which adds to the degree of confidence one may place on the loss development pattern for Multiple Peril, and most other short-tailed lines, is the low amount of first year reserves that are in the form of IBNR. IBNR, which stands for "Incurred But Not Reported," refers to all the losses that the insurance company has theoretically had on a line of business but have not yet been reported to the company. These claims will eventually be paid, after the loss has been reported, so the insurance company must reserve against them. The smaller the amount of reserves that are in the form of IBNR, the more confidence the company actuaries have that they have accounted for all eventual losses. Generally, only 15% of Multiple Peril losses are not reported before the end of the first year.

Figure 7

Multiple Peril
Loss Development



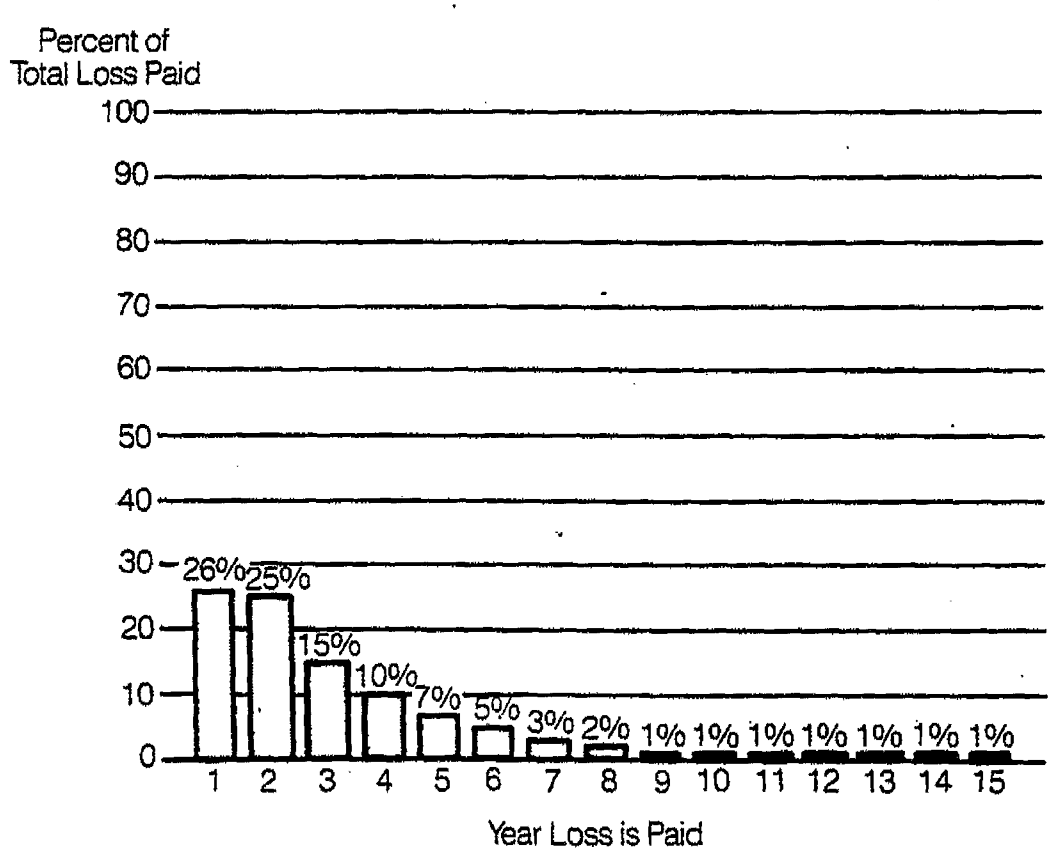
Workers' Compensation

Another form of liability insurance is Workers' Compensation, which provides benefits for employees injured on the job, or who incur a job-connected disease. Unlike traditional liability coverages, Workers' Compensation pays benefits without regard to whether the employer is at fault. The worker can be paid without establishing negligence on the part of the employer.

Workers' Compensation is the classic example of a long-tailed line of insurance. This long tail arises from two circumstances: 1) many claims take years to develop, and 2) many loss claims are paid in the form of recurring medical cost payments or income supplements. Occupational diseases, such as asbestosis, can take decades to reach their disabling stage. Even with legal questions aside as to which insurer is responsible for loss coverage (either the company that was the insurer when the disease was contracted or the one when the symptoms appeared), the loss reserves for Workers' Compensation extend well beyond 10 years. The long tail and the high degree of confidence that can be placed on certain portions of Workers' Compensation losses, such as disability income payments, have caused this line to become one of the few lines of P&C insurance for which discounted reserve accounting is used. The topic of discounting reserves is covered in Section VII of this report.

Figure 8

Workers' Compensation Loss Development



General Liability

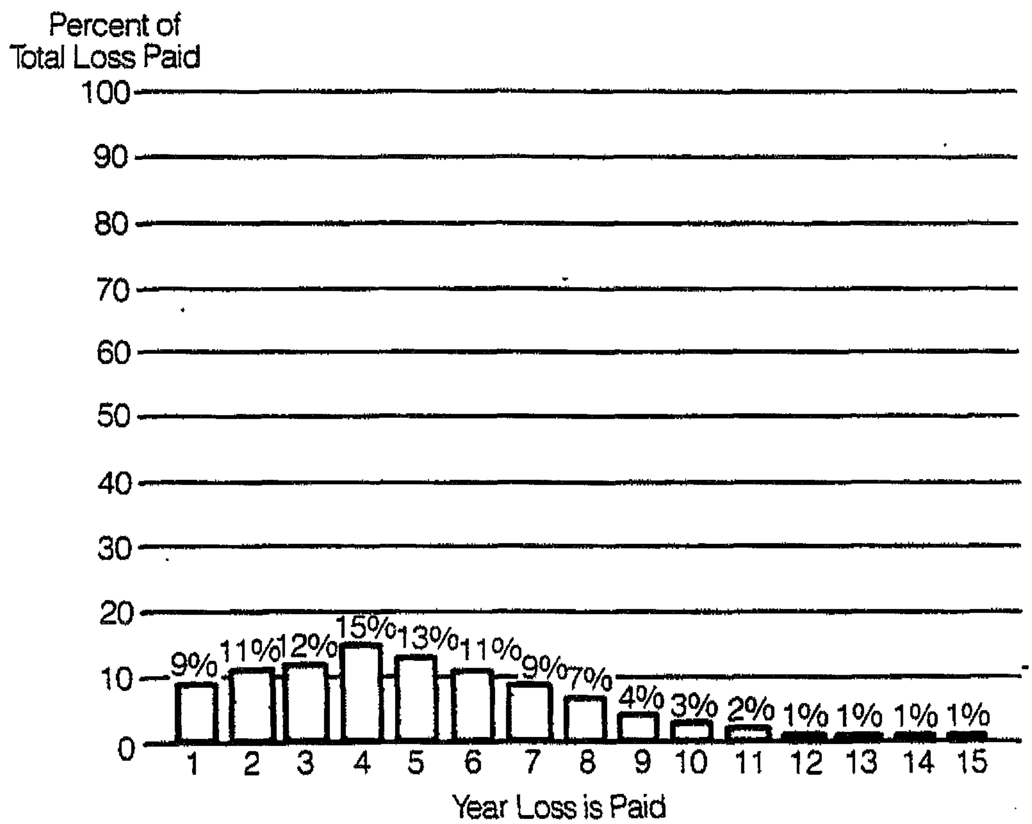
Of all the lines of insurance, General Liability, which includes Medical Malpractice in our examples, poses the most difficult problem in estimating the loss development pattern and size of loss reserve. Public attitudes toward responsibility on such issues as product and toxic substance liability are evolving toward a definition of strict liability and court awards are increasing at an alarming rate. The actuary who is standing in 1985 and is trying to develop loss reserves for claims that will be paid as a result of a medical malpractice trial in 1990 is, at best, using a cloudy crystal ball. Best's analysts state that General Liability reserves for losses arising in 1978 have had to be increased 25% through year-end 1983, and it is estimated that over 40% of these losses are still yet to be paid![8]

The loss development pattern used in Figure 9 is meant only to be representative of the patterns that develop for General and Medical Malpractice Liability. This pattern indicates the lag before most cases come to trial which causes an increasing pattern in losses in years one through four. The very long tail can result from years of litigation and, finally, settlement in the form of many years of support payments.

Figure 9

General Liability

Loss Development



8. "1983 Casualty Loss Reserves," op. cit.

The Liabilities in Aggregate

Once the individual lines of business have been analyzed, the next step is to aggregate them into a representation of future liabilities of the entire company. Figure 10 shows the projected loss payments for XYZ Casualty for the business written in 1972 (that is, as of year-end 1972). To construct Figure 10, the various loss payment patterns are simply weighted by the dollar amount of loss reserves that are held for each line at year-end. The *loss reserves* themselves are a function of the amount of premiums that were earned and what the expected loss ratio is for that line of insurance. The *loss ratio* for any line of insurance is the amount of premium, stated as a percentage, that is expected to be ultimately paid out in the form of a loss, including the cost of adjusting that loss. There is a counterpart to the loss ratio which is known as the expense ratio. The *expense ratio* is the portion of each premium that is used by the insurance company to pay for the costs of writing insurance—overhead, sales compensation, etc. While the expenses may, in theory, also have a development pattern associated with them, we will assume that all expenses are paid in the year that a premium is earned.

Figure 10

XYZ Casualty
Liability Schedule

1972

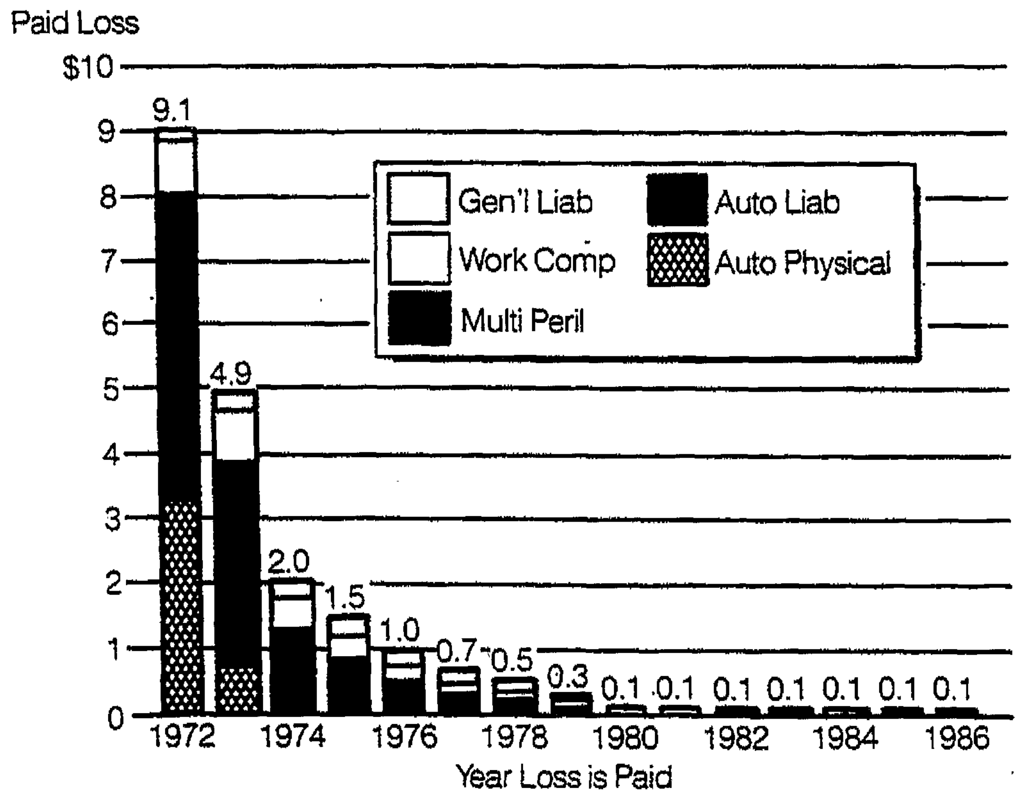
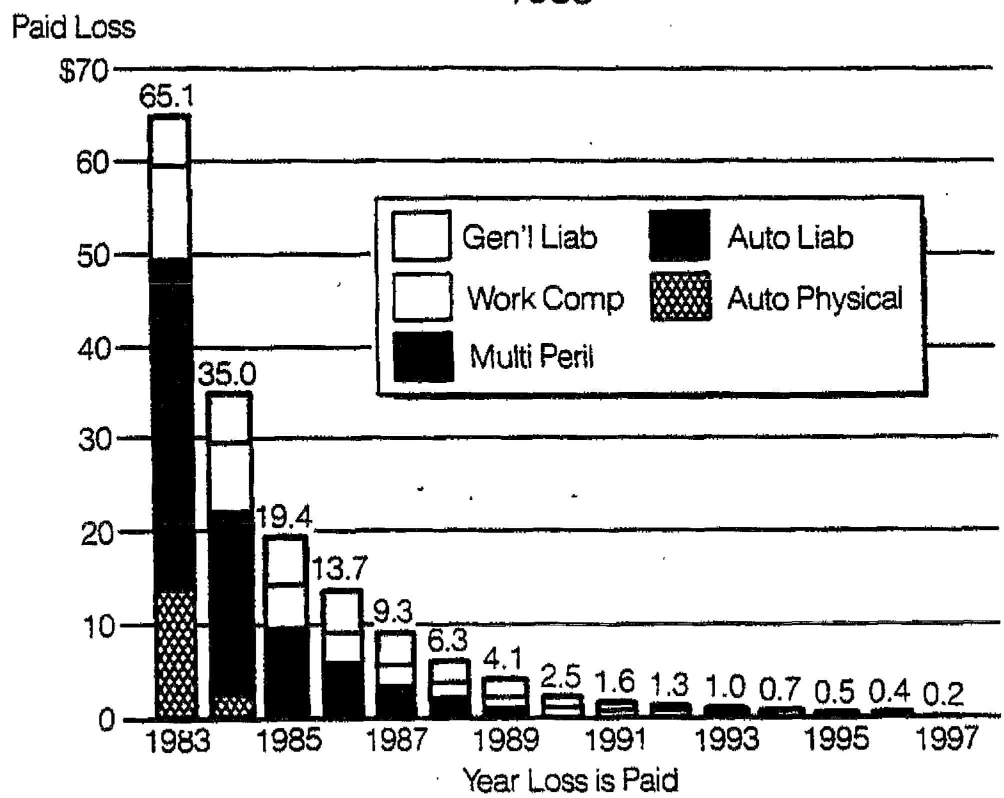


Figure 10 represents the paid-loss profile for a company that has been in existence for only one year. As time goes by, the liability profile changes, including a component from business that was written in the most recent year and a component from losses that have not yet been paid on previous years' business. At year-end 1983, XYZ Casualty has 11 years of underwriting experience. This mature company represented in Figure 11 has approached a paid-loss profile that will exhibit very little change from year-to-year, impacted only by changes in the mix of business, pricing of future coverage, and losses developing differently from assumptions as to incidence and claim size. It is against this liability profile that we must construct an investment strategy that will help support and increase the total market value of the company.

Figure 11

XYZ Casualty
Mature Liability Schedule

1983



SECTION III

MARKET VALUE OF A PROPERTY & CASUALTY INSURANCE COMPANY

The total market value of any firm is derived from two sources: portfolio equity and franchise equity.[9] Portfolio equity equals the value of *currently* booked assets less the value of *currently* booked liabilities. The value of net income from business *not yet booked* by the insurance company is termed franchise equity and is equivalent to the economically determined value for goodwill. While a proper asset/liability model should consider both forms of equity in the measurement of risk exposure. the practical problems in measuring franchise equity preclude it from most analyses. This paper will continue the traditional emphasis on measuring the interest rate risk exposure of portfolio equity only. Portfolio equity. as proxy for the total value of the firm. does however represent an excellent target account for interest rate risk management.

The size of portfolio equity for a P&C insurance company depends upon the difference between the cash flows from currently booked assets (its investment account) and the amounts needed to pay insurance claims on currently booked coverages (its loss profile). There are a number of useful ways portfolio equity can be analyzed, the simplest being a visual representation of all cash flows in the year they are to be received or paid. Just as we aggregated the liabilities of a P&C insurance company into a profile (Figure 11). we can aggregate its assets. Figure 12 displays this representation of assets for XYZ Casualty as of 12/31/83. Asset flows are shown by layering the coupon and maturity cash flows from each of the investment portfolio's holdings on top of one another. Short-term money-market instruments (T-Bills) are shown as a lump sum in 1984. their maturity year. as are the firm's holdings in common stock. Because XYZ's investment policy of purchasing 20-year bonds has been in effect for only the past 11 years. we can see that the bond maturity ladder does not commence until 1992. The ladder. however. is then continuous from 1992 through 2003. An even more mature P&C company will have this ladder entirely filled in. but the large amount of growth the industry experienced in the 1970s indicates this asset profile is probably not unlike many insurers today.

The liabilities are put in the same context by placing the paid loss profile (developed in Figure 11) over the asset profile. Figure 12 now represents all of XYZ Casualty's tangible accounts: assets. liabilities. and *portfolio equity* (the amount assets exceed liabilities).

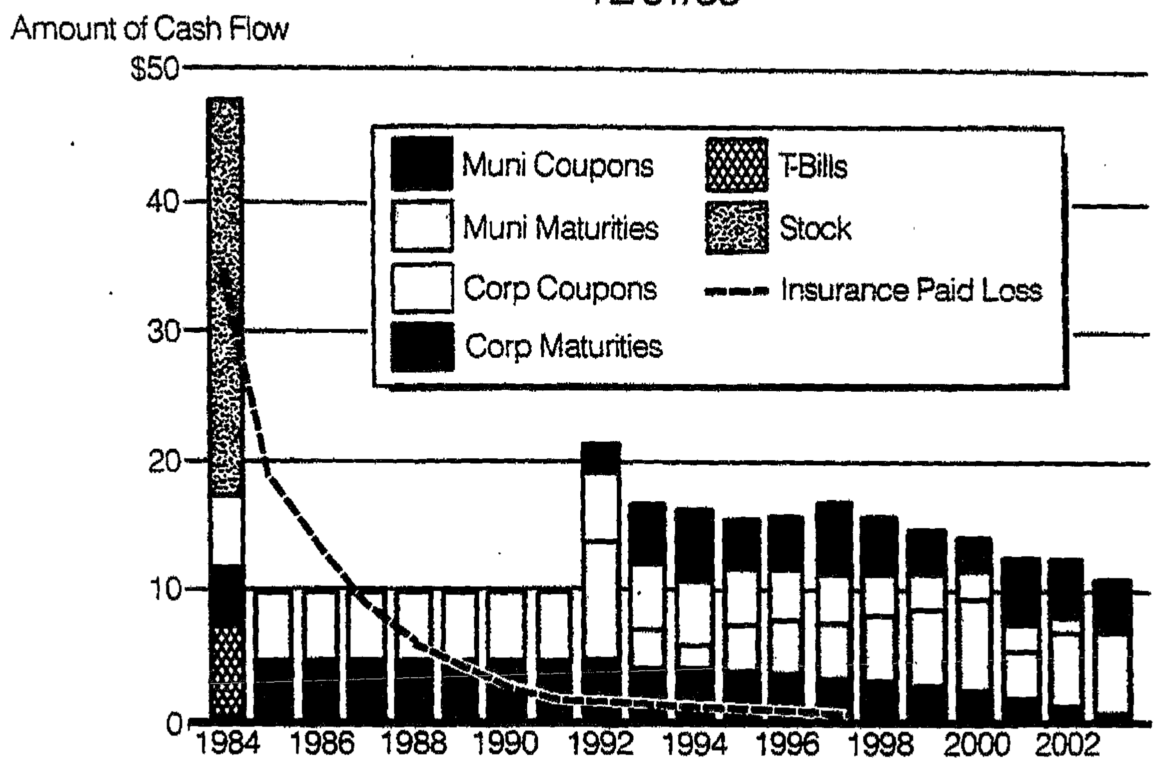
How should one value portfolio equity? Even though assets clearly exceed liabilities in every year except 1985 and 1986, the firm's portfolio equity may or may not be positive. Should we "borrow" some excess assets from later years to plug the gap and then simply sum up the excess assets? Or should we give some weight to the fact that the shortfalls occur sooner than most of the excesses and derive some "present value" concept of portfolio equity by allowing for the time value of money? The latter is clearly the preferred method. but even this answer is somewhat ambiguous because there are many ways to discount cash flows.

9 The concepts concerning the market value of equity described in this section are covered more fully for other financial institutions by Toevs and Haney in "Measuring and Managing Interest Rate Risk: A Guide to Asset Liability Models Used in Banks and Thrifts," Morgan Stanley, October 1984

Figure 12

XYZ Casualty
Assets and Liabilities

12/31/83



Book Value Accounting: Statutory Surplus

Book value accounting would have us discount most asset cash flows at the yields at which they were *originally* purchased. Stocks would be shown at their current market value (as would any assets in default). On the other hand, because the P&C industry does not utilize discounting of reserves, the firm's liabilities as represented by the paid loss profile are, in effect, discounted at 0% interest (that is, the losses are simply summed).[10] The portfolio equity value that results from this calculation (Table III) is known as Statutory Surplus and is the same as shown previously in Table I.

Current Value Accounting: Current Value Surplus

Some have argued that book value accounting is inappropriate for companies that must compete in a market value world. Since 1981 the Progressive Corporation (an Ohio P&C insurance company) has included in its annual report to shareholders financial statements on both a GAAP[11] basis and a basis they term "Current Value/Total Return" accounting; "current value" referring to values on the balance sheet and "total return" referring to values on the income statement.

10. Some discounting of P&C reserves is becoming more prevalent, particularly in Workers' Compensation, but this practice has been ignored for the illustrative purposes of this paper.

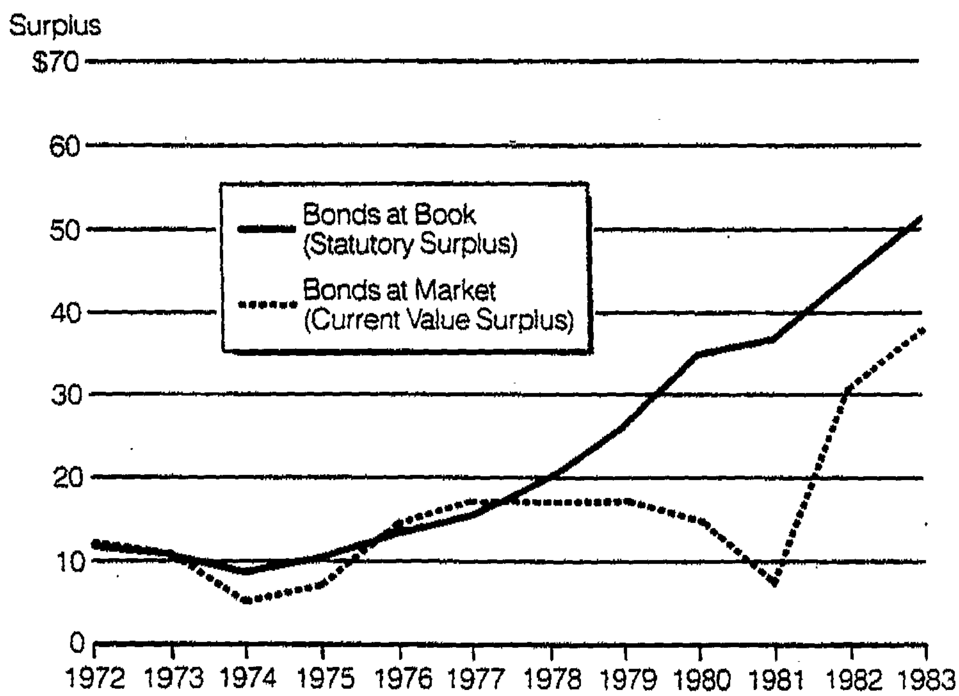
11. Generally Accepted Accounting Principles (GAAP) for P&C companies is similar to Statutory Accounting in its treatment of the valuation of financial assets. GAAP, however, is on an accrual basis for some expenses and some reserves.

Progressive Corp.'s methodology is to value all assets at current market values and to include any unrealized capital gains or losses (reduced by the long-term capital gains tax rate) in both the balance sheet and the income statement. Neither statutory nor GAAP accounting practice recognizes unrealized gains or losses on book value assets, and unrealized gains or losses on stocks are shown only on the balance sheet (as an adjustment to surplus or shareholders' equity). Progressive states that "total return" accounting "most accurately reflects real economic results."[12] The Current Value Surplus (CVS) for XYZ Casualty is also developed in Table III.

Progressive's current value/total return accounting system certainly would have highlighted the plight of the P&C industry as interest rates rose during 1974-75 and then again beginning in 1979. Figure 13 shows the surplus of XYZ Casualty on both a statutory basis (with bonds held at book value) and a current value basis (with bonds at market) for the period from 1972 to 1983. While statutory surplus exhibits a somewhat constant rate of positive growth, with absolute declines in only 1973 and 1974 caused by the bear market in stocks, the CVS displays quite an erratic path. Current value accounting would have shown the industry to have dangerously low amounts of surplus in 1974 and 1981, amounts that were only 61% and 20%, respectively, of reported statutory surplus. Even with the bond rally of 1982-83, the CVS of XYZ Casualty is only 74% of statutory surplus at year-end 1983.

Figure 13

XYZ Casualty
Statutory Surplus
vs. Current Value Surplus



12. Progressive Corp., 1981 Annual Report; page 3.

Table III-A

Statutory Surplus (12/31/83)

	Description	Discounting Used?	Discounting Rate	Value
Assets:				
	Money Market	Yes	Cost	$ 7,353
	Taxable Bonds	Yes	Cost	50,350
	Tax-exempt Bonds	Yes	Cost	58,822
	Stock	N/A	Market	30,529
	Total Assets			$147,054
Liabilities:				
	Loss Reserves	No	N/A	$ 95,960
Statutory Surplus:				$ 51,094

Table III-B

Current Value Surplus (12/31/83)

	Description	Discounting Used?	Discounting Rate	Value
Assets:				
	Money Market	Yes	Market	$ 7,353
	Taxable Bonds	Yes	Market	43,765
	Tax-exempt Bonds	Yes	Market	52,056
	Stock	N/A	Market	30,529
	Total Assets			$133,703
Liabilities:				
	Loss Reserves	No	N/A	$ 95,960
Current Value Surplus:				$ 37,743

Table III-C

Market Value Surplus (12/31/83)

	Description	Discounting Used?	Discounting Rate	Value
Assets:				
	Money Market	Yes	Market	$ 7,353
	Taxable Bonds	Yes	Market	43,765
	Tax-exempt Bonds	Yes	Market	52,056
	Stock	N/A	Market	30,529
	Total Assets			$133,703
Liabilities:				
	Loss Reserves	Yes	After-Tax	$ 78,631
Market Value Surplus:				$ 55,072

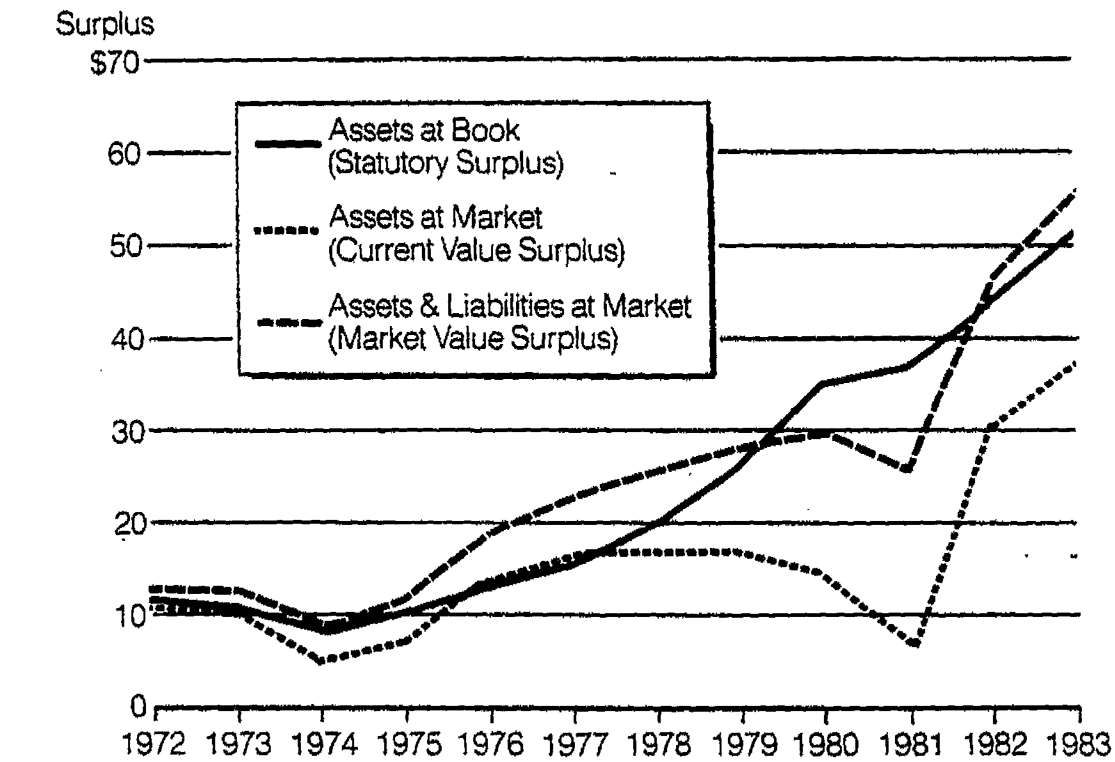
Economic Accounting: Market Value Surplus

In reality, current value accounting goes too far in one direction in placing a market value on the firm's surplus account. While we have put a market value on all of the firm's financial *assets*, we have still ignored the time value of money in valuing the firm's *liabilities*. In a true economic sense, any losses that are to be paid in the future should be discounted in a manner similar to the way we discounted future cash flows on assets. Any positive rate of interest used in discounting these reserves would, of course, lower the present value of the loss reserve and, thus, increase the reported amount of surplus. Questions arise as to what rate of interest should be used in discounting P&C reserves. Possibilities are: the firm's historical yield on its asset portfolio (a recommendation of the U.S. General Accounting Office); the yields for bonds in the current investment market; a conservative rate of interest, such as 4%, to account for the potential variation in loss experience; or no discounting at all (statutory accounting). For purposes of this paper, we will use the spot yields for municipal bonds in the current investment market as a proxy for the after-tax investment yield available on current investments. The result of placing a market value on both assets and liabilities, Market Value Surplus (MVS) for XYZ Casualty at year-end 1983, is also shown in Table III.

The first thing that strikes us about XYZ's year-end 1983 MVS is that it is *higher* than the reported statutory surplus ($55,072 versus $51,094). Does this mean that XYZ is healthier

Figure 14

XYZ Casualty
Surplus Values



than the reported statements indicate? It is possible. If XYZ has accurately projected its loss payment pattern (a big "if" in today's environment), and we have used the appropriate after-tax rates of discount, then we can truly say that the insurer has a current MVS which is higher than the reported statutory surplus. But, as shown below in Figure 14, the MVS of XYZ has also been volatile, almost as volatile as CVS. In addition, even though the application of a discount rate to an insurer's liabilities will always cause MVS to be higher than CVS, and often higher than statutory surplus, there have been periods (1980 and 1981) when MVS has been lower than statutory surplus. If the industry continues its traditional investment practice of investing in long-term bonds, MVS will continue to be highly volatile. There exists, however, an alternative. Company management can choose to *manage* the volatility of MVS and help insure the profitable growth of the firm.

SECTION IV

MANAGING MARKET VALUE SURPLUS

Why should an insurance company manage its MVS, as opposed to managing statutory surplus or no management at all? One obvious reason is that a firm's MVS is a leading indicator of the future book value of the firm, since book value converges to market value as asset and liability items mature. Managing the MVS will thus enable company management to better serve the goals of the company's owners—that is, higher future stock prices for stockholders or higher future dividends for policyholders as the case may be.

Another reason for managing market value surplus may come from insurance regulators who, from time to time, have become concerned with the capital adequacy of the industry. A 1979 report commissioned by the National Association of Insurance Commissioners (NAIC) had the actuarial consulting firm of Tillinghast, Nelson & Warren (TNW) review the appropriateness of using various forms of market value accounting, including the consideration of some type of Mandatory Securities Valuation Reserve (MSVR) for P&C company balance sheets.[13] While TNW's ultimate recommendation was that an MSVR is not appropriate for P&C insurers, they did make a number of observations as to the appropriate size and valuation of an insurer's surplus account, concluding that statutory accounting is appropriate only if asset/liability matching is adhered to:

> *...Thus, we believe much more attention must be paid to the matching of asset and liability maturities. If liabilities are reasonably stated, and if matching is reasonably carried out, then amortized values will not result in significant surplus distortions.*[14]

> *...The required surplus for insurers who keep their asset and liability maturity schedules in balance with each other should be less than for those who choose to invest in longer maturities, thereby accepting the additional risk of market price fluctuations.*[15]

Interest Rate Sensitivity of Surplus: Duration Gap

To understand how to manage the value of a firm's MVS, one must first understand what can make that value change. The MVS of an insurer can be viewed as a "net" bond, albeit one with an unusual series of positive and negative cash flows. These flows are represented by the amounts assets differ from liabilities, developed previously in Figure 12 and shown below in Figure 15. Like any bond, the MVS will have a price sensitivity to changes in interest rates—the "longer" the bond is, the more sensitive its price will be to a given change in interest rates. The particular measure of "length" that accurately communicates the interest rate sensitivity of the "net bond" is known as *duration.*[16]

---

13. "Appropriateness of a Mandatory Securities Valuation Reserve Concept for Property and Casualty Insurers": Parts 1 & 2: Tillinghast, Nelson & Warren: February 1978 and March 1979.

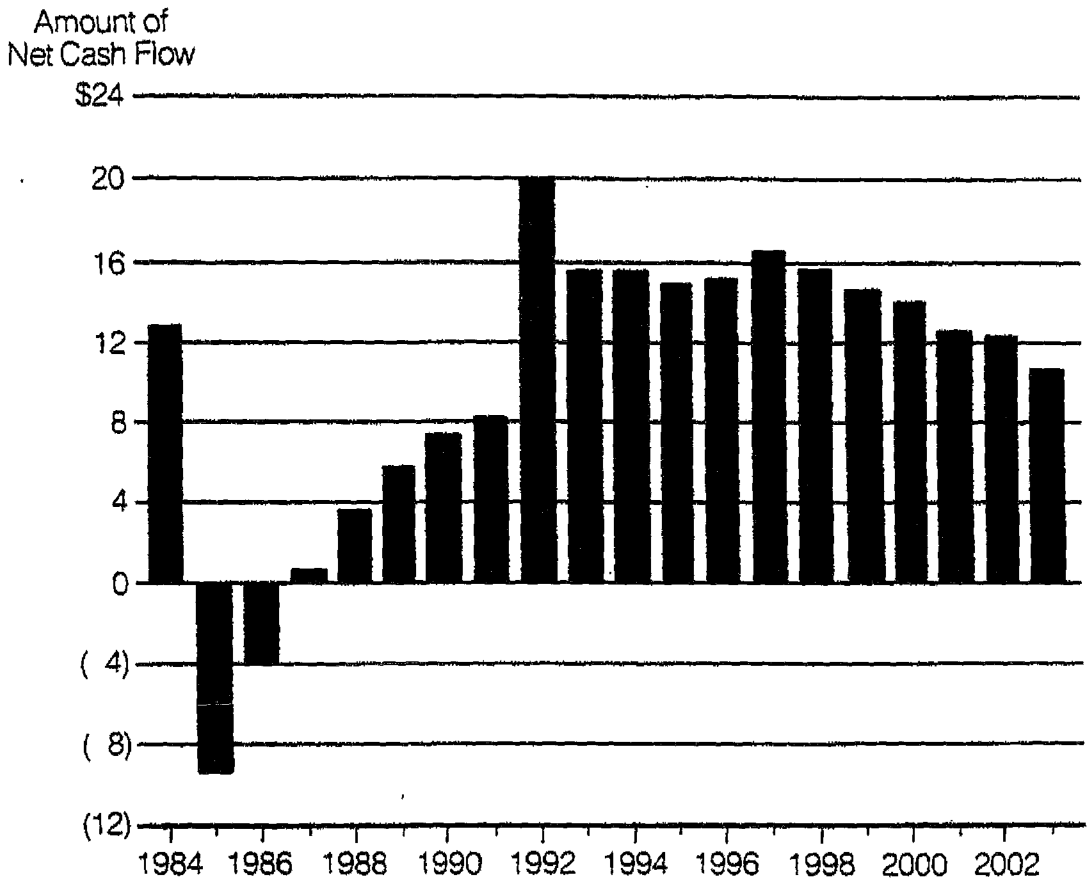
14. TNW report: part 1; page 68.

15. TNW report: part 1; page 101.

16. This paper presupposes a familiarity with the concept of duration. For a more fully developed explanation of duration, see Toevs (Jan., 1984) and, for the particular context of P&C insurers, "Duration," an article by Ronald Ferguson in the Proceedings of the Casualty Actuarial Society, volume 70 (1983).

Figure 15

XYZ Casualty
Cash Flow of Net Surplus



The calculations required to find, and then manage, the duration of the MVS account are based upon the additive property of duration—the duration of a combination of instruments is equal to the sum of the durations of the individual instruments weighted by their appropriate market values. Table III had shown us that the MVS of an insurer is equal to the market value of its assets (MV_A) minus the market value of its liabilities (MV_L), or:

$$MVS = MV_A - MV_L \qquad (1)$$

Using Equation 1, we can weight each component's duration (D) by its market value (MV):

$$D_{mvs} * MVS = (D_{mva} * MV_A) - (D_{mvl} * MV_L) \qquad (2)$$

Therefore, to find the duration of MVS:

$$D_{mvs} = \frac{(D_{mva} * MV_A) - (D_{mvl} * MV_L)}{MVS} \qquad (3)$$

The amount that the duration of MVS varies from zero is known as the firm's "duration gap."[17] Insurers with a larger duration gap will have a market value surplus that is more susceptible to changes in interest rates than firms with a gap of zero. A positive duration gap, indicating that assets are longer than liabilities, would mean that any rise in interest rates would lower the absolute value of MVS (since asset market values would decline relatively more than liability market values). A negative duration gap would mean that rising interest rates would actually improve the amount of MVS.

Table IV develops the duration gap for XYZ Casualty as of year-end 1983. The durations for the fixed-income assets (bonds and money-markets) are calculated using the traditional method of weighing their various cash flows' terms-to-maturity by their present values, this method is sometimes referred to as Macaulay's duration. The common stock component of XYZ's asset base presents an interesting duration calculation problem, since common stock does not have a maturity date or maturity value. It can be shown that the duration of a consol bond (a bond which pays coupons into perpetuity) is approximately 1/i, where i is its yield to maturity. Similarly, the duration of common stock may be approximated by 1/d, where d is the current dividend rate.[18] In our model, we assume d to equal the annual dividend yield on the S&P 500. Even though we can calculate a duration for stock, the value of stock is influenced by many other factors in addition to interest rates, indicating that the duration value for stock is, at best, an approximation for its sensitivity to interest rates. Section IV of this paper will show how a company's stock holdings can be modified to make them fit better into an asset/liability management framework.

Table IV

Market Value Surplus (12/31/83)

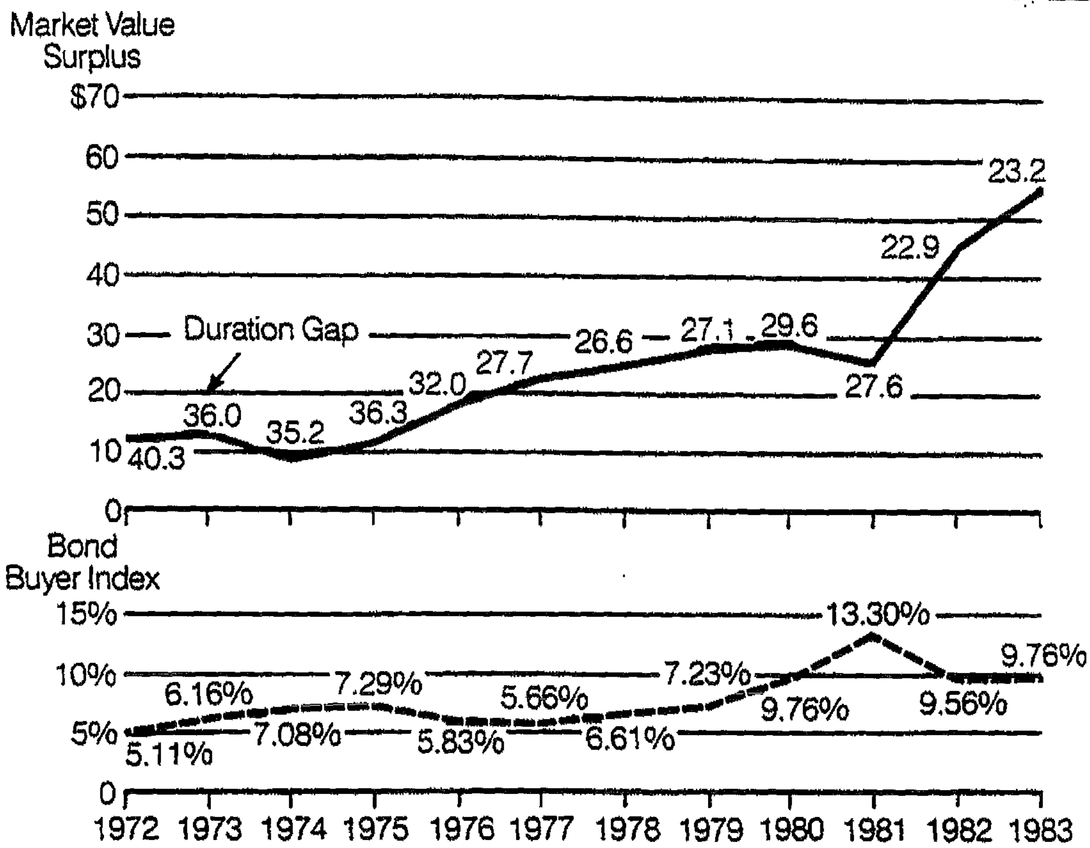
	Description	Discounting Used?	Discounting Rate	Value	Duration
Assets:					
	Money Market	Yes	Market	$ 7,353	1.00 yrs
	Taxable Bonds	Yes	Market	43,765	7.33
	Tax-exempt Bonds	Yes	Market	52,056	8.41
	Stock	N/A	Market	30,529	23.26
	Total Assets			$133,703	11.04 yrs
Liabilities:					
	Loss Reserves	Yes	After-Tax	$ 78,631	2.51 yrs
Market Value Surplus:				$ 55,072	23.22 yrs

17. The concept of Duration Gap is more fully developed in Toevs and Haney (1984).

18. Attempts have been made to show the duration of equity securities to be a function of the security's Beta. The reader is referred to Bierwag, Kaufman and Toevs (1983) for studies relating to the duration of equity securities.

Figure 16

XYZ Casualty
Duration Gap of MVS



XYZ's positive duration gap of 23.2 years indicates that its MVS will have the interest rate sensitivity equal to that of a 23.2 year zero-coupon bond (which is a very long term bond); that is, a rise in interest rates will cause a large decline in MVS but any decline in interest rates will sharply increase MVS. Figure 16, which takes the line for XYZ's MVS from Figure 14 and shows it in relation to movements in interest rates (the Bond Buyer Index), indicates the effects of carrying a large positive duration gap over a long period of time. We can see that MVS has been quite volatile, falling precipitously in the high interest rate environments of 1974 and 1981, and rebounding dramatically with the drop in interest rates in 1982.

Duration Gap of Target Accounts

Rather than just blindly carry a large duration gap, which will leave the value of the firm's MVS greatly dependent upon the level of interest rates (an uncontrollable factor), the insurer should choose to manage this gap. By deciding when, and to what extent, to have a duration gap, company management can position the firm to take advantage of any projected changes in interest rates, or, in an extreme case, to position the firm to have a MVS that is immune to changes in interest rates. The previous section gave us the tools that are necessary to compute the duration gap, this section will give us a few targets at which a company can shoot.

Duration Gap of Surplus

The duration gap previously computed in Equation (3) and Table IV can more specifically be called the Duration Gap of Surplus (DG_s), in that it tells the interest sensitivity of an insurer's market value surplus account. As explained previously, a large positive DG_s indicates a potentially volatile surplus account, one that will increase (decrease) in value if interest rates should decline (rise). A natural target for managing DG_s would be to achieve a DG_s that is equal to zero, which would mean that the value of MVS will be *immune* to changes in interest rates.

Having a DG_s of zero would be valuable to most insurers, since insurance regulators typically limit premium volume to a small multiple of reported surplus. If the surplus is managed so that its value never declines from changes in interest rates, then premium volume will not have to be constricted, unless large underwriting losses develop. Setting DG_s equal to zero would give the MVS account the stable price characteristic of a money-market fund (which also has a duration near zero). However, like a money-market fund, a DGs of zero would result in fluctuations in total earnings, depending upon the interest rate environment.

Duration Gap of Total Return on Surplus

In reality, it is probably unduly restrictive, *and unnecessary*, for an insurer to take its duration gap of surplus all the way down to zero. Another valuable feature of duration is that it indicates the holding period over which a rate of return can be immunized. If duration is set equal to the desired holding period, then the initial promised return can be realized independent of changes in the level of interest rates.[19] This would suggest that by setting a firm's DG_s equal to some holding period, the net yield of MVS can be achieved over that period. Many insurers desire to manage *annual* returns-on-surplus so that they are always positive. Provided the net yield of MVS is positive, this goal can be achieved by setting DG_s equal to one.

Immunizing a total rate of return over longer planning horizons can be achieved by having a higher DG_s. Even though a higher DG_s immunizes MVS over a longer holding period, interim results may be above or below the immunized holding-period return. The insurer must reach some compromise between holding-period return and stability in interim results. In general, the formula for the Duration Gap of Total Return-on-Surplus (DG_{trs}) is:

$$DG_{trs} = Ds - H \qquad (4)$$

Where D_s is the duration of surplus and H is the holding or investment period over which management wishes to lock up the currently available return on surplus. By setting $DG_{trs} = 0$, the firm will immunize total return over the holding period.

19. Provided the duration is rebalanced periodically to counteract duration drift.

Duration Gap of Leverage

A final target account is of interest to those who are concerned with capital adequacy. It is the ratio of MVS to MVA. The reciprocal of the MVS to MVA ratio is the MVA to MVS ratio, known as "economic leverage." The economic leverage of a firm will remain unchanged only when both elements either remain unchanged or change proportionately. That is, when interest rates cause the market value of assets to increase by one percent, the market value of surplus must also increase by one percent for the asset-to-surplus ratio to remain unchanged. Since the percentage change for the market value of any security (given a change in interest rates) depends upon its duration, the duration of surplus must equal the duration of the assets to immunize economic leverage against interest rate fluctuations. Therefore, to immunize the economic leverage of a firm, it is necessary to set the Duration Gap for Economic Leverage (DG_{el}) equal to zero, where:

$$DG_{el} = D_{mvs} - D_{mva} \tag{5}$$

SECTION V

DEVELOPING AN INVESTMENT STRATEGY

The purpose of this section of the paper is to illustrate the development and implementation of an investment strategy. This strategy is developed using the duration gap techniques discussed in Section IV. We then test the strategy, in Section VI, over the same historical period used in our previous examples. The results of the illustrative investment strategy are finally compared to XYZ's results that were achieved using a more traditional investment strategy.

Management Goals

Before constructing any investment strategy, it is necessary for company management (usually the Investment Committee of the Board of Directors) to enumerate the various goals and policies that will govern the Investment Department's operations. Once these goals and policies have been outlined, an investment strategy can be created that will optimize the results while staying within policy guidelines. The goals we have listed below are not meant as a recommendation, each company's needs and policies are unique, but these goals are indicative of the types of goals management may voice as well as being reasonable *and achievable:*

1) *Achieve consistently a positive growth rate in shareholders' investment.* Alternative ways of stating this goal are growth in: shareholder's equity; book value per share; policyholders' surplus; or market value surplus. The goal of *positive* growth is necessary to insure that available growth in premium sales is never needlessly impaired by a decline in statutory surplus. *Consistent* growth should translate into a higher price/earning multiple for the firm's stock, as the market tends to place a premium on consistency.

2) *Risk should not be taken with policyholder funds—investment of insurance reserves should be as "risk neutral" as possible.* Insurance regulators are concerned with the adequacy of both the insurer's reserves for insurance claims and the assets that back those reserves. Changes in the investment environment should not impact the insurer's ability to pay claims as they come due. Since shareholders' ownership interest, which is protected by goal #1, is equal to the portion of the company that is not needed for insurance reserves, goals #1 and #2 are complimentary.

3) *Maximum flexibility should be maintained in shifting between income from taxable or tax-exempt investment sources.* Sudden changes in insurance tax laws or the profitability of insurance underwriting may make investment income from taxable, rather than tax-exempt, sources more desirable. Another way of stating this goal is to maximize after-tax income on a year by year basis.

4) *Consistent with the foregoing goals, maximize the firm's investment exposure to equity securities.* To illustrate other types of goals that may be achieved, we will make the assumption that XYZ's management believes that the equity markets will outperform fixed-income securities over a long time horizon. This extra performance may be needed to achieve the real rates of return that are necessary to cover insurance claims that may grow through the ravages of unexpected inflation.

Implications of Goals on Investment Strategy

Achieve consistently a positive growth rate in shareholders' investment. This goal implies that the Market Value Surplus (MVS) of the firm should never decline on an absolute basis. If the relevant time horizon for the firm is the next annual report to shareholders (although that may not be the most optimal time horizon), then "never decline" can be restated to mean "MVS at year-end will at least be equal to the previous year's MVS."

There are basically two ways to insure that MVS never declines: 1) manage MVS as if it were a one-year, or shorter, bond; or 2) purchase a "put" option on MVS that has a strike price equal to the previous year's MVS. Section IV showed that by managing the Duration Gap of Surplus (DG_s) or the Duration Gap of Total Return on Surplus (DG_{trs}) to a value of zero, the MVS will act either like a money-market account ($DG_s = 0$) or will have a rate of return that is immunized over the one-year time horizon (if $DG_{trs} = 0$ for $H = 1$). Provided the net yield of MVS is positive, managing the DG_s or DG_{trs} to a value of zero will maintain a positive growth in MVS. Because a DG_{trs} of zero is equivalent to a DG_s of one[20], and immunizing any holding period shorter than one year will also achieve a positive yearly growth in MVS, we can relax the constraint that DG_s equal zero to DG_s can be between zero and one, inclusive, or:

$$0 \leq DG_s \leq 1 \qquad (6)$$

While it is not possible to purchase directly a put on the MVS account, it is possible to purchase puts on many of the components of MVS. If the DG_s is managed so that it is less than the one-year time horizon, then put protection is theoretically unnecessary, since downside risk of MVS return has already been eliminated by immunizing its holding period return. But, as mentioned in Section III, the common stock component of MVS cannot be managed entirely in a duration framework. Stocks do sometimes fall in value when we expect them to rise! However, by utilizing a "protective put" strategy, stocks can be used in an investment strategy that is consistent with the goal of never having a decline in MVS. The "protective put" strategy will be explained later in this section.

Risk should not be taken with policyholder funds—investment of insurance reserves should be as "risk neutral" as possible. The funds that are set aside to pay future insurance claims must be sufficient no matter what economic environment develops. The simplest way to achieve sufficiency is to make sure that there is enough cash flow from bond maturities and coupon income to cover the paid loss profile developed in Section II. The assets that are needed to generate the necessary cash flow would be earmarked for "expense and insurance reserves" and would not be available for any investment practice that deviates from an asset/liability match. This method is known as "cash flow matching" or "bond dedication" in investment management terminology. Cash flow matching works because there is never a possibility of taking a loss on an asset if assets never have to be sold.

An alternative method of achieving the "risk neutral" position is to make sure that there is always sufficient *market value* of assets available to pay any insurance claims. This implies that it is also necessary that if the present value of the assumed claims should vary, then the market value of the assets must vary to the same degree. Assuming that a change in interest rates is the only factor that can change the value of the firm's assets, the assets' *duration* will

20. Since $DG_{trs} = D_s - H$, where H is a one-year holding period, and $DG_s = D_s$, then $DG_s = 1$ if $DG_{trs} = 0$.

convey the amount of that value change. Initially, a sufficient market value of assets is earmarked for expense and insurance reserves. If the duration of the assets is equal to the duration of the insurance claims and expenses, then there will always be a sufficient market value of assets to cover claims. This investment management technique is known as "duration matching" and it generates results that are close to those of cash flow matching.[21] The primary cause of any difference in results between the two techniques would be due to yield curve twists — cash flow matching is fully immune to twists, but duration matching techniques may have results that vary. However, duration matching is often more cost-effective than cash flow matching, since duration-matched portfolios can utilize the entire universe of available investments while cash-matched portfolios can select from only the investments that have the appropriate cash flow characteristics.

Maximum flexibility should be maintained in shifting between income from taxable or tax-exempt investment sources. Insurance companies always have the option of switching their investment portfolio between taxable and tax-exempt investments, but it is seldom viable if the switch would require the realization of a capital loss (and, thus, a decline in statutory surplus). To achieve this goal of flexibility the insurer must either have investments that never decline in value, or a sufficient cash flow from asset maturities to effect the desired rebalancing through new purchases. Investments that never decline in value are few and far between, but the previously mentioned strategy of holding a stock portfolio with "protective puts" would fit the bill. The alternative of sufficient maturity cash flow could be achieved by simply holding an asset portfolio with a relatively short average life, but this might be in conflict with goals #1 and #2, which call for duration matching the longer liabilities.

Consistent with the foregoing goals, maximize the firm's investment exposure to equity securities. Since simply holding a large portfolio of equities would subject the firm's surplus account to potential losses, should stock prices decline, this method would be inconsistent with the first goal of constant positive growth in surplus. The "protective put" strategy is an answer to this potential problem. While a protective put truncates any downside risk in holding equities, it still allows the company to participate in much of the upside potential of the equity market.

A Strategy

While one might suspect that the four goals of company management could be somewhat mutually exclusive (i.e., "you can't have your cake and eat it too"), there exists a relatively simple investment strategy that more than adequately achieves the targets in the framework of our model environment. The following three steps are all that are required:

1) Duration-match the paid loss profile.
2) Invest 100% of statutory surplus in common stocks with protective puts.
3) Invest the balance of assets, if any, in short term instruments.

21. *In fact, cash flow matching is merely a polar case of duration matching in which not only durations have been matched, but cash flows as well*

Duration-Match the Paid Loss Profile

By setting aside enough assets to cover the paid loss profile, we not only protect the policyholder, but we give ourselves a "free hand" to do whatever we want with the balance of the assets. Since the duration of XYZ Casualty's paid loss profile is only 2.51 years as of 12/31/83 (Table VI), and the market value of the loss reserves is approximately 60% of total asset market value, it is quickly apparent that this strategy requires purchasing many more short and intermediate term assets than the typical P&C insurer is accustomed to buying (a duration of 2.51 years is equivalent to a maturity average life of approximately 3 years). While some investment managers might perceive that shortening the bond portfolio so dramatically could result in a reduction in investment income (particularly in tax-exempt bonds, where there is typically a steep, positively sloped yield curve), this should not be a major concern. Once we have taken care of the protection of policyholders, we are free to more aggressively posture the remainder of the asset portfolio to go after higher rates of return (with commensurately higher degrees of risk). It is possible that the incremental return gained on the balance of the portfolio may more than offset the perceived income give-up of the shortened bond portfolio.

Stocks with Protective Puts

The idea of purchasing common stocks and simultaneously purchasing a put option on that stock has been given many labels. Some choose to call the strategy "portfolio insurance" to connote that the put option provides "stop loss" protection against a decline in stock prices.[22] Many portfolio managers may already be familiar with related strategies, sometimes called "90-10" or "money-market/options" portfolios.[23]

This relation comes about through the operation of "put/call parity" which says that owning a put on a stock is the same as owning a package of securities composed of:

1) A long position in a call option.
2) A short position in the underlying stock.
3) A sufficient amount of cash that will mature to the exercise price of the option if invested at the risk free rate of interest.

Intuitively, one can see that this package is equivalent to a put option since:

—if stock prices fall, the call expires worthless and, after the shorted stock is delivered, we are left with the difference between the exercise price (the cash portfolio) and the stock price (which is less than the cash portfolio).

—if stock prices rise, we must deliver the shorted stock which has appreciated. Since we are holding cash whose value is the exercise price, and the call option will pay off any appreciation above the exercise price, we have a net return of zero.

Since a protective put is purchased against an existing stock position, a put-protected portfolio is equal to:

$$\text{Put protected portfolio} = \text{Put} + \text{Stock} \qquad (7)$$

22. See Platt and Latainer (1984), and Rubenstein and Leland (1981).

23. For example, Fredman and Moore, "The Money Market/Options Portfolio: A Prudent Strategy." *Best's Review*. 9/82.

And we have shown that a put is equivalent to a portfolio of a call, cash and shorted stock:

$$\text{Put} = \text{Call} + \text{Cash} - \text{Stock} \tag{8}$$

Then a put-protected portfolio also equals a position in cash plus a call option:

$$\begin{aligned}\text{Put protected portfolio} &= \text{Call} + \text{Cash} - \text{Stock} + \text{Stock}\\ &= \text{Call} + \text{Cash}\end{aligned} \tag{9}$$

This equivalency of a protective put being equal to holding a cash portfolio plus a call option has given rise to the terminology of "90/10" (that is, invest 90% in cash and 10% in call options) or "money-market/options" portfolio, even though the latter strategies are only crude approximations to a correctly executed protective put strategy.

Many insurance companies are restricted from outright purchases of put options, but the preceeding paragraphs show that this does not encumber the company in executing a protective put strategy, since the company can replicate the position by holding money-market instruments and purchasing call options. In addition, it is also possible to replicate the *call options*, through dynamically rebalancing a portfolio composed of money-market instruments and the underlying stock. The proof of this equivalency is rather involved, and the interested reader is referred to other work on the subject[24], but the beauty of this second equivalency is that it means that the insurance company can hold protective puts against *any* security in its portfolio, not just securities that have options that are traded on listed or over-the-counter exchanges. Indeed, these arguments can be expanded to include replication of options for entire *portfolios* of securities, especially applicable to an institutional investor such as an insurance company.

Remainder in Short Term Instruments

This part of the strategy is rather self explanatory, since most insurance companies are already accustomed to holding large pools of liquid, short term, securities. But these short term instruments do not necessarily have to be the traditional money-market investments used by most insurers (e.g., T-Bills, bank CDs and BAs, commercial paper, repurchase agreements, institutional money-market funds). The short term component of the portfolio can be composed of any instrument that is liquid (readily marketable) and has the duration characteristics of short term securities (stable principal value). Other, less traditional, investments include: municipal commercial paper (if the insurer needs tax-exempt income); preferred stocks (which are often an attractive alternative to tax-exempt commercial paper); option conversions (a technique consisting of long positions in a put option and stock and a short position in a call option, that yields a risk free rate of return); and interest rate swaps (which can be used to shorten an asset's duration). Even the money-market position taken in the process of replicating the protective put strategy, explained previously, can use these alternative short term investments.

Another interesting way to view the portfolio's short term component is described in a product Morgan Stanley calls "cash management".[25] This product measures the tradeoff between added risk and added return from any investment strategy. These tradeoffs help inform cash

24. See Rubenstein and Leland, op. cit.

25. See Haney, "A Risk Controlled Approach to Managing Corporate Cash Pools," Morgan Stanley & Co. Incorporated, forthcoming.

managers of their probable investment performance. For any large cash pool that is likely to remain intact for many years (as is the case for a healthy percentage of an insurer's "liquidity pool"), a strategy of simply rolling-over short term investments will not have much principal risk, but there will be a substantial amount of reinvestment risk. The cash management product gives the portfolio manager a measure of the risk involved in deviating from the shortest investment strategy and may allow some incremental return to be prudently captured by extending the portfolio's duration.

Are the Goals Achieved?

How well does the illustrated investment strategy perform in achieving company management's four goals? Goal #2 (risk neutral investment of insurance reserves) is achieved through a duration matched portfolio. This portfolio's assets will always be sufficient to cover any projected paid losses.[26] The short average life of the assets invested behind the loss reserves, together with the large component of assets held in short term investments, makes goal #3 (maximum tax flexibility) very achievable. By utilizing a protective put strategy in conjunction with holding common stocks, the percentage of assets devoted to stocks can be prudently increased and goal #4 (maximum exposure to equity securities) can be attained. The final, and most important goal, #1 (positive growth in shareholders' investment), should be achieved by the very short duration gap of surplus the foregoing strategy implies (if the downside risk of holding equities has been truncated, and the balance of the surplus funds are invested in instruments with short durations, then the duration gap of surplus will be short).

It appears that the suggested investment strategy should perform well, but "seeing is believing" and an example helps to illuminate the points. Section VI tests the suggested strategy on our model company—XYZ Casualty.

26. The potential problem of paid losses being different from projections is covered in Section VII of this report.

SECTION VI

TESTING THE INVESTMENT STRATEGY

The method used to test the illustrated investment strategy is to use the model company constructed in Section I (XYZ Casualty) and replace its "traditional" investment strategy with our developed one. By using the same economic time period (1972-1983) and not tampering with the basic sales and liabilities of XYZ (i.e., the amount and mix of premiums and the loss profiles for claims paid are identical to Section I), we can see in isolation what effect changing the investment strategy will have on the reported, as well as market value, surplus accounts.

Assumptions

A few simplifying assumptions are made to facilitate the handling of: 1) XYZ's changing tax position, 2) any problems associated with modelling actual sales of investments and 3) any dynamic strategy of asset allocation (such as would occur in a protective put strategy). These assumptions are:

1) The percent of assets invested in Municipal securities is held constant at 40%.
2) Each year's loss profile is immunized through *cash flow matching*. Since we are assuming that losses develop as projected, cash flow matching will eliminate the need to sell any assets. Both taxable and tax-exempt investments required for the cash flow matching strategy are developed from historical yield curves and are indexed so that the 20-year part of the curve is equal to either the Bond Buyer Index or Moody's Corporate Bond Composite Index.
3) The annual return for the equity portion of the portfolio is derived from studies that Morgan Stanley has done regarding equity portfolios with protective puts. In particular, the protective put strategy that is used insures a floor return of zero in conjunction with investing in the S&P 500. All returns from the portfolio insurance strategy are given net of the transactions costs involved in rebalancing the strategy's money-market and equity portfolios. Since insurance companies have a wide range of flexibility in deciding when to actually *realize* any appreciation in their stock holdings, the model assumes that any appreciation that is in excess of the annually available return on the risk-free asset (T-Bills) is kept unrealized. This has the effect of deferring payment of taxes to later periods but also lowers *reported* investment income in years where stocks greatly appreciate (Statutory and GAAP accounting do not show unrealized gains or losses on the income statement). Furthermore, it is assumed that *all* income derived from the stocks with protective puts is taxed at *ordinary* income rates. In actual practice some of the return will be derived from dividends and, possibly, long term capital gains, both of which are tax preference items. We felt it is the most conservative approach to evaluate the suggested investment strategy using the worst possible tax rates.

Table V shows the yields that were historically available for duration matched portfolios versus the longer, 20-year, investments in which XYZ had been investing. The table also shows the annual returns that could be achieved in an equity portfolio with protective puts versus the returns available on the non-protected S&P 500.

Table V

New Investment Strategy

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Bond Portfolio												
Duration of Purchases	2.51	2.61	2.57	2.58	2.66	2.69	2.66	2.63	2.54	2.42	2.54	2.51
Yields												
Taxable Purchases	7.43%	7.66%	9.01%	8.21%	6.33%	7.70%	10.59%	12.66%	15.61%	14.83%	11.39%	11.59%
(vs 20-Yr Yield)	7.47	8.05	9.63	9.57	8.47	8.54	9.49	11.35	14.04	15.38	13.02	13.01
BP Pick-up (Give-up)	(4)	(39)	(62)	(136)	(214)	(84)	110	131	157	(55)	(163)	(142)
Municipal Purchases	3.53%	4.21%	5.30%	4.89%	3.47%	4.09%	5.81%	6.48%	8.17%	9.29%	6.45%	6.95%
(vs 20-Yr Yield)	5.11	5.16	7.08	7.29	5.83	5.66	6.61	7.23	9.76	13.30	9.56	9.76
BP Pick-up (Give-up)	(158)	(95)	(178)	(240)	(236)	(157)	(80)	(75)	(159)	(401)	(311)	(281)
Stocks with Protective Puts												
Total Annual Return	N/A	-0.01%	-0.01%	23.74%	15.59%	0.13%	2.87%	16.39%	27.90%	0.04%	12.74%	18.94%
(assumed realized)	N/A	-0.01	-0.01	7.30	6.30	0.13	2.87	12.90	12.20	0.04	12.74	8.80
(assumed unrealized)	N/A	0.00	0.00	16.44	9.29	0.00	0.00	3.49	15.70	0.00	0.00	10.14
S&P 500 Total Return	N/A	-14.86%	-25.25%	34.15%	22.07%	-7.20%	6.48%	18.48%	33.38%	-5.58%	20.79%	22.17%

Results

The new investment strategy would have performed quite admirably over the previous 11 years. The annual income statements and balance sheets that emerge under the new strategy are reproduced in Table VI so that the reader can make any comparison to XYZ's previous results (Table II) of greatest interest. We have chosen to highlight two comparisons between the new strategy and the old: 1) the growth and stability of the various surplus accounts (statutory, current value and market value), and 2) the amount of annual investable cash flow (which would enable flexibility in tax management).

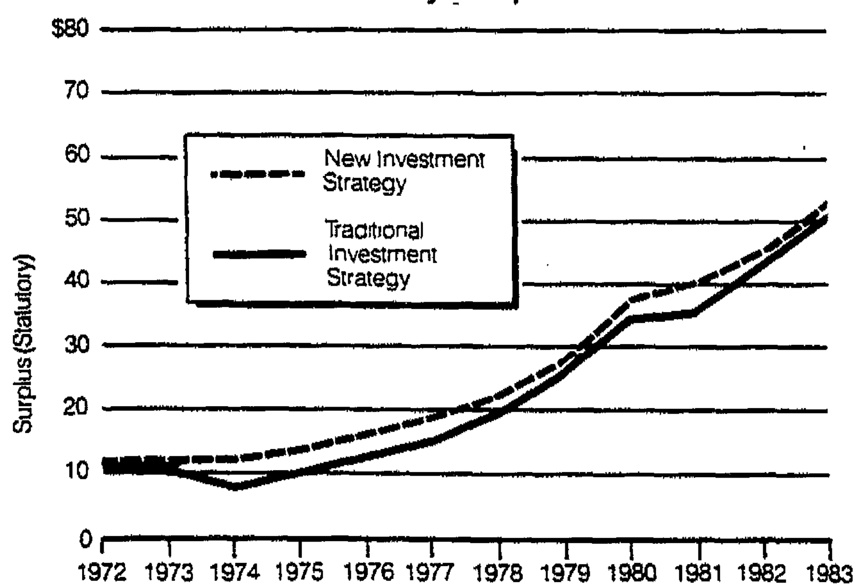
Figure 17 shows the development of Statutory, Current Value and Market Value Surplus for both our new investment strategy and for XYZ's historical results (which are taken as a proxy for actual industry results). Statutory surplus develops on a very similar path to the old investment strategy; however, because of the protective put strategy used in the new investment strategy's stock portfolio, statutory surplus did not decline noticeably in either 1973 or 1974 (as opposed to the large declines under the old investment policy). The new strategy results in a very stable growth in statutory surplus.

The current value surplus also exhibits very stable growth under the new investment strategy. With the relatively short average life of the bond investment portfolio, together with the protective put on the stock portfolio, the run-up in interest rates in 1975 and 1979-81 did not cause any substantial market depreciation in the new strategy's assets. This can be contrasted to the horrendous current value surplus results achieved under the old investment strategy.

Figure 17

New Investment Results Compared to Traditional

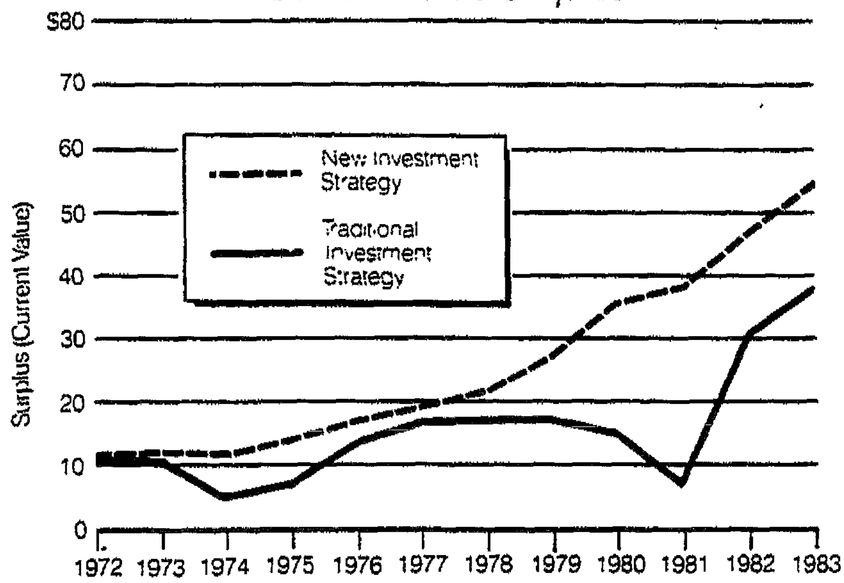
Statutory Surplus



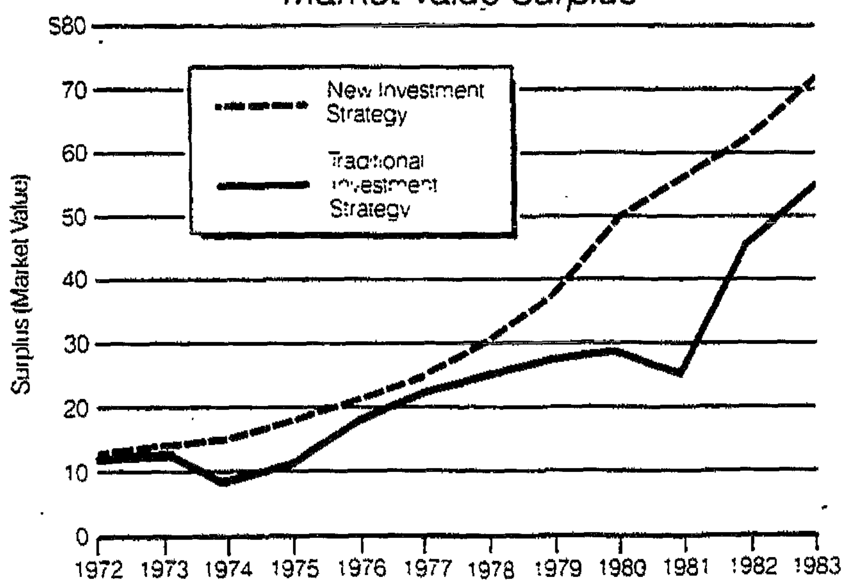
Finally, market value surplus shows a dramatic improvement over XYZ's results (which were, as we pointed out in Section III, already above statutory results). The new investment strategy outperforms the old strategy on *both a reported and an economic basis*. Furthermore, the new results could be expected to outperform the old strategy even under different economic environments because we have chosen to partake in *structured* risk taking, not *market* risk taking. Because of the large, positive, duration gap of surplus maintained in the old strategy, the old strategy would outperform the new strategy only if the markets behaved in our favor (i.e., interest rates declined and then stayed low, or were stable for a long period of time). The new strategy should perform well in all economic environments.

Figure 17 (continued)

Current Value Surplus



Market Value Surplus



As for the real-life world where a company has to worry about such things as taxes, Figure 18 shows that the new investment strategy presents the added flexibility of a large amount of annual cash flow. High cash flow is sometimes thought of as a negative by bond portfolio managers, since there exists the risk that the cash must be reinvested in a lower interest rate environment, but our "risk neutral" investment strategy has eliminated that concern. High cash flow now means high flexibility.

Figure 18

Investable Cash Flow

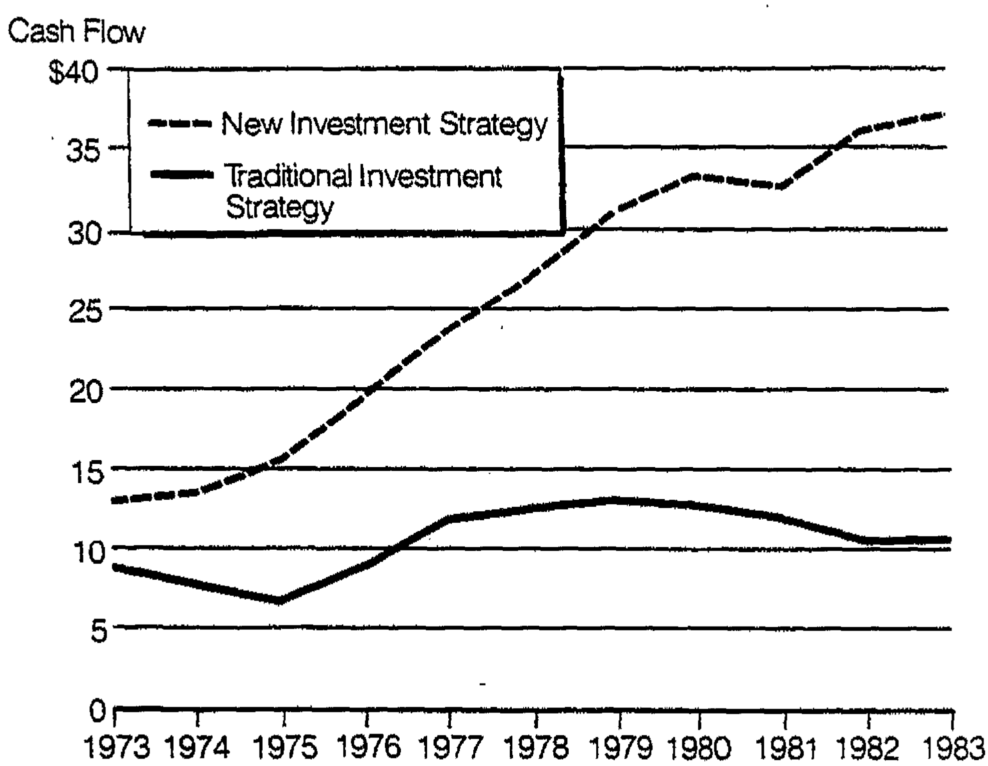


Table VI

XYZ Casualty
With New Investment Strategy

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Income Statement												
Net Premium Earned												
Auto Liab	$11,383	$11,744	$11,889	$12,807	$15,122	$18,159	$20,073	$21,682	$22,934	$24,011	$25,610	$27,528
Auto Phys	6,408	6,864	7,106	7,399	8,962	10,961	12,484	13,969	15,466	16,469	17,579	19,178
Workers Comp	4,010	4,676	5,332	5,965	7,274	9,072	10,771	12,653	13,895	14,407	13,918	14,097
Multi Peril	5,083	5,941	6,815	7,751	9,328	11,309	13,162	15,082	16,544	17,562	18,828	20,014
General Liab	2,564	2,734	2,921	3,644	5,038	6,571	7,380	7,726	7,797	7,370	7,077	7,240
Total Premium	$29,448	$31,959	$34,063	$37,566	$45,724	$56,072	$63,870	$71,112	$76,636	$79,819	$83,012	$88,057
Expenses												
Losses Incurred	20,519	22,974	26,201	30,712	35,610	40,415	45,489	52,542	57,794	62,401	67,338	72,846
Underwriting Exp	7,842	8,613	9,339	9,862	11,355	13,546	15,763	17,696	19,430	20,915	22,412	24,069
Net Underwriting Income	$ 1,087	$ 372	($ 1,477)	($ 3,008)	($ 1,240)	$ 2,111	$ 2,618	$ 874	($ 588)	($ 3,498)	($ 6,738)	($ 8,858)
Investment Income												
Money Market	0	69	177	279	315	282	502	1,174	1,386	2,004	2,481	1,593
Taxable	0	82	334	600	764	877	1,116	1,433	1,708	1,727	1,888	1,963
Tax-Exempt	0	324	505	716	902	966	1,173	1,623	2,153	3,068	3,826	3,916
Stocks (Realized)	0	(1)	(1)	874	878	21	545	2,209	3,434	15	5,193	4,039
Investment Exp	0	(76)	(103)	(126)	(155)	(188)	(223)	(261)	(308)	(367)	(402)	(442)
Pre-tax Investment Income	$ 0	$ 398	$ 912	$ 2,343	$ 2,705	$ 1,959	$ 3,114	$ 6,878	$ 8,373	$ 6,447	$12,986	$11,068
Pre-tax Operating Income	$ 1,087	$ 770	$ (565)	($ 665)	$ 1,464	$ 4,070	$ 5,732	$ 7,752	$ 7,785	$ 2,950	$ 6,248	$ 2,210
Income Taxes	522	214	(514)	(663)	270	1,490	2,188	2,942	2,590	(54)	1,114	(785)
Net Operating Income	$ 565	$ 556	($ 51)	($ 2)	$1,194	$2,580	$3,544	$4,810	$5,194	$3,004	$5,134	$2,995

Table VI

XYZ Casualty
With New Investment Strategy

(continued)

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Balance Sheet												
Assets												
Money Market	$ 1,185	$ 2,397	$ 3,827	$ 5,001	$ 5,754	$ 7,064	$ 9,104	$11,358	$14,115	$17,231	$18,102	$19,634
Taxable	1,108	4,352	7,083	9,216	11,945	14,451	15,786	16,465	14,805	15,094	16,458	16,525
Tax-Exempt	9,172	12,513	15,252	18,768	22,748	27,012	31,625	37,311	44,450	48,724	53,636	59,801
Common & Pfd	11,465	12,020	11,969	13,934	16,423	19,003	22,547	28,144	37,756	40,760	45,894	53,543
Total Assets	$22,929	$31,282	$38,131	$46,920	$56,871	$67,530	$79,062	$93,279	$111,126	$121,810	$134,090	$149,502
Liabilities & Surplus												
Loss Reserves	$11,465	$19,262	$26,162	$32,986	$40,447	$48,527	$56,515	$65,135	$73,370	$81,049	$88,196	$95,960
Surplus	11,465	12,020	11,969	13,934	16,423	19,003	22,547	28,144	37,756	40,760	45,894	53,543
Total Liab. & Surplus	$22,929	$31,282	$38,131	$46,920	$56,871	$67,530	$79,062	$93,279	$111,126	$121,810	$134,090	$149,502
Ratio of Market Value to Book:												
Taxable Bonds	100%	100%	99%	100%	102%	100%	97%	96%	92%	95%	101%	100%
Tax-Exempt Bonds	100%	100%	98%	100%	102%	101%	99%	98%	97%	96%	102%	101%

SECTION VII

MISCELLANEOUS ISSUES

This final section touches upon a few other issues that P&C insurance companies are dealing with today. Future Morgan Stanley publications will expand upon many of these issues from both an analytical and an applications viewpoint, but their existence must at least be noted for this publication to be complete.

Unexpected Loss Development

The problem of losses ultimately being paid either in a different amount or at a different time than original expectations is a very real one. If losses are larger or sooner than expected, the amount of assets set aside for Market Value Reserves will not grow (with investment income) to a sufficient amount to cover the losses. Since the illustrated investment strategy recommends that any assets not needed for reserves be invested in higher risk instruments, it is very important that the reserves allow for any contingencies.

Large pools of losses developing sooner than expected, while significantly lowering the amount of investment income that can be earned on reserves, is probably not a major cause of underreserving. While there should be a tendancy for insureds to attempt to expedite filing for claims when interest rates are high, it is unlikely that insurers feel much pressure to speed their already fast processing of claims. An acceleration of individual, highly uncertain, claims may be a factor at some insurers, but it is probably not an industry-wide phenomenon. The more likely cause of any underreserving is growth in the *ultimate size* of the losses. Many property claims are specifically indexed to inflation, such as replacement cost coverages, and liability claims are susceptible to increase through their own rate of inflation. This paper's reserving and investing strategies have not yet overtly accounted for any unexpected inflation rate, but that does not mean they cannot be adapted to account for one.

The duration matching investment strategies, mentioned previously, immunize a *nominal* return over some holding period; that is, targets are based upon an expected dollar payment of claims and an investment portfolio is then constructed that will pay off at least that much in market value. But the process of estimating the amount of these *expected* claims incorporates only an expected rate of inflation. *Unanticipated* inflation may cause these reserves to be insufficient. What the insurer ideally would like to immunize is a *real* rate of return—one that pays off in "units of coverage" or some similar concept. Even if one could determine the correct rate of inflation to which returns should be indexed, there still exists the problem that traditional fixed-income instruments, such as bonds, have flows that are denominated in fixed dollars. Unless there is some mechanism for these flows to increase with inflation, duration matching strategies will not immunize a real rate of return.[27]

One method of having investment returns keep pace with inflation is to purchase investments which have inflation dependent returns. Real assets, such as real estate, and common stocks come to mind when discussing the topic of assets which might keep pace with inflation. While it is beyond the scope of this paper to quantify whether these assets do indeed yield real rates of

27. Indexed bonds, where either the coupons or principal value change with inflation, may be issued at some time in the future and would be a useful instrument in immunizing real rates of return. See Babbel (1984).

return (there has been much debate on the subject), the previously mentioned problems of these assets' volatility make their use as the core holding for an insurer's reserve position a difficult proposition.

A second method of keeping investment returns current with inflation is to have investments that roll over quite often. If yields for similar investments in the new market environment are related to the inflation rate, then rolling the portfolio into these new investments will achieve the inflated return. The problem with this strategy is that it requires a departure from the duration matching principle, since shorter maturities are required if the portfolio is to be kept current.

A far simpler method of assuring sufficient assets for inflated claims is to intentionally *overestimate* the expected size or rapidity of the losses. The amount of the overestimation could be earmarked as a *contingency reserve* and would be invested in the same manner as the market value of the estimated loss reserve. The American Academy of Actuaries and the Society of Actuaries have recommended this approach in their report on the role of a life insurance company's valuation actuary.[28]

The joint committee's recommendation is to apportion the assets of the insurance company among:

1) Valuation Reserves—these reserves would be sufficient under expected circumstances, but there is still some probability that additional reserves are required.
2) Contingency Surplus—this is the amount of surplus (assets minus valuation reserves) that is required to bring the probability of ruin down to an acceptably low level.
3) Vitality Surplus—this remaining portion of assets is the amount that is available for growth and change. It is the vitality surplus that we would recommend be used for investment strategies with higher degrees of risk.

Table VII gives an idea how large the contingency surplus of a typical P&C insurance company could be, given some assumptions as to ultimate size and timing of claim payments. The *expected* reserve requirement, "valuation reserve" in the preceding categorization, is found at the intersection of "Change in mean life of loss profile" of zero years and "Overall actual-to-expected ratio for claim amounts" of one times (shown by the shaded regions). This amount, $78,631, is the same amount that was used in Section III to calculate the firm's Market Value Surplus. If the insurer wished to hold assets in reserve that are sufficient to cover claims that are *both* 20% higher than expected and occur, on average, one year sooner than expected. Table VII shows that reserves in the amount of $100,574 would be required (bold print). This contingency surplus would increase needed reserves to 127.9% of the reserve amount calculated under a straight expectations approach and would be 104.8% of the reserve calculated under statutory accounting methods. The stair-step line in the middle section of the table indicates that the increased reserve would also be sufficient to cover other combinations of claim size and incidence. For example, the new reserve would be sufficient to cover claims that total 1.5 times the expected amount, provided they are paid 3 years later than expected.

28. "Final Report of the Joint Committee on the Role of the Valuation Actuary in the United States," American Academy of Actuaries and Society of Actuaries, February 1985.

In actual practice, as in our suggested investment strategy, much of the hypothetical contingency surplus is already being held in the form of short-term assets. These short-term assets not only provide for catastrophic claims, but they also should provide for some protection against unanticipated inflation as they are rolled over in differing interest rate environments.

Table VII

Required Reserves for Unanticipated Loss Development

Market Value of Required Assets

				Overall Actual-to-Expected Ratio for Claim Amounts							
		Mean Life	Asset Duration Target	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Change	−2 yrs	0.93	0.88	$71,755	$80,724	**$89,694**	$98,663	$107,632	$116,602	$125,571	$134,540
In	−1 yrs	1.93	1.74	$67,049	$75,431	**$83,812**	$92,193	**$100,574**	$108,955	$117,336	$125,718
Mean	0 yrs	2.93	2.51	$62,905	$70,768	$78,631	$86,494	$ 94,357	$102,220	$110,084	$117,947
Life	+1 yrs	3.93	3.22	$59,222	$66,625	$74,027	$81,430	$ 88,833	$ 96,235	$103,638	$111,041
Of	+2 yrs	4.93	3.88	$55,923	$62,913	$69,903	$76,893	$ 83,884	$ 90,874	$ 97,864	$104,855
Loss	+3 yrs	5.93	4.50	$52,946	$59,565	$66,183	$72,801	$ 79,420	$ 86,038	$ 92,656	$ 99,274
Profile	+4 yrs	6.93	5.08	$50,245	$56,526	$62,806	$69,087	$ 75,368	$ 81,648	$ 87,929	$ 94,210

As A Percent of Market Value Reserves

				Overall Actual-to-Expected Ratio for Claim Amounts							
		Mean Life	Asset Duration Target	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Change	−2 yrs	0.93	0.88	91.3%	102.7%	114.1%	125.5%	136.9%	148.3%	159.7%	171.1%
In	−1 yrs	1.93	1.74	85.3%	95.9%	106.6%	117.2%	**127.9%**	138.6%	149.2%	159.9%
Mean	0 yrs	2.93	2.51	80.0%	90.0%	100.0%	110.0%	120.0%	130.0%	140.0%	150.0%
Life	+1 yrs	3.93	3.22	75.3%	84.7%	94.1%	103.6%	113.0%	122.4%	131.8%	141.2%
Of	+2 yrs	4.93	3.88	71.1%	80.0%	88.9%	97.8%	106.7%	115.6%	124.5%	133.4%
Loss	+3 yrs	5.93	4.50	67.3%	75.8%	84.2%	92.6%	101.0%	109.4%	117.8%	126.3%
Profile	+4 yrs	6.93	5.08	63.9%	71.9%	79.9%	87.9%	95.9%	103.8%	111.8%	119.8%

As A Percent of Statutory Reserves

				Overall Actual-to-Expected Ratio for Claim Amounts							
		Mean Life	Asset Duration Target	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Change	−2 yrs	0.93	0.88	74.8%	84.1%	93.5%	102.8%	112.2%	121.5%	130.9%	140.2%
In	−1 yrs	1.93	1.74	69.9%	78.6%	87.3%	96.1%	**104.8%**	113.5%	122.3%	131.0%
Mean	0 yrs	2.93	2.51	65.6%	73.7%	81.9%	90.1%	93.3%	106.5%	114.7%	122.9%
Life	+1 yrs	3.93	3.22	61.7%	69.4%	77.1%	84.9%	92.6%	100.3%	108.0%	115.7%
Of	+2 yrs	4.93	3.88	58.3%	65.6%	72.8%	80.1%	87.4%	94.7%	102.0%	109.3%
Loss	+3 yrs	5.93	4.50	55.2%	62.1%	69.0%	75.9%	82.8%	89.7%	96.6%	103.5%
Profile	+4 yrs	6.93	5.08	52.4%	58.9%	65.5%	72.0%	78.5%	85.1%	91.6%	98.2%

Discounting of Loss Reserves

The subject of discounting reserves, as well as reserve sufficiency, has recently received much attention. The United States General Accounting Office (GAO) and Treasury have both made proposals to utilize some form of discounting in reporting of insurance reserves for tax accounting purposes. If a lower, discounted, reserve is used for reporting purposes, then more income would be reported at the time the insurance is written. Increased reported income would result in increased tax expense, a natural goal of both departments.

Many observers note that the P&C insurance industry is potentially already severely underreserved and any discounting that is used will increase their tax liability and only exacerbate a bad situation.[29] While this paper recommends utilizing discounting of reserves in the *management* of assets and liabilities, it does not contemplate any additional tax burden that discounting would impose.

Cash Flow Underwriting

To most industry followers, "cash flow underwriting" is a term used to indicate that an insurer is involved in the practice of deliberately pricing insurance coverage so that it has a statutory loss (that is, its combined ratio is over 100). The rationale for writing this business is the hope that the claim will not have to be paid until investment income sufficient to cover the loss has been earned. An alternative definition was advanced by Warren Buffet in Berkshire Hathaway's 1980 Annual Report in which he suggested that some companies were going beyond pricing practices that are based upon future investment income. These companies, he says, are pricing based upon the need to generate sufficient premium income to *avoid selling* any book value assets (bonds) at current market values which would generate book losses. He categorizes this type of cash flow underwriting as "asset maintenance" underwriting.[30]

The methods we have suggested in this paper may, to some readers, be a recommendation for some form of cash flow underwriting. While it is true that our methods recognize the cash flow from future investment income, they differ from traditional cash flow underwriting in one very important way—*we take risks on only a structured basis*. It is one thing to simply take risk, as in cash flow underwriting—it is entirely another thing to understand what risk you are taking. The asset/liability methodology described in this paper accommodates even the most risk-averse insurance company management.

How Do We Get There From Here?

Once the P&C insurer has decided to adopt the methods and strategies suggested in this paper (if it has not already done so), there remains the question of how to restructure the asset portfolio. Because the combination of statutory accounting practice and an environment of historically high interest rates has led to a situation in which most insurers have assets with market values that are substantially below book values, most insurers cannot simply, painlessly, swap their predominately longer assets for the assets we have suggested. It is a

29. "Discounting the Downtrodden," *Forbes*, February 25, 1985.

30. Berkshire Hathaway Corp., 1980 Annual Report.

shame that this situation exists, since it is statutory accounting which now hinders the economic health of an industry it was designed to protect.

There are, however, many methods of diminishing the reported accounting effects of any asset restructuring. *Interest rate swaps* may be used to alter an asset's duration while, effectively, amortizing the imbedded economic loss over a longer time horizon. Selling bonds or preferred stocks in conjunction with *issuing puts*, which give the purchaser the right to sell the asset back to the insurance company, are being used as a method of delaying the accounting impact of a market loss. In addition, many *swaps* of assets may have such a quick payback period, or such a large economic return for a given amount of reported loss, that the taking of a reported loss is justified.

CONCLUSION

We hope that this publication has been instructive. By building a simplified model of the P&C insurance industry, we have been able to illustrate the ebbs and flows of insurance underwriting and investment returns. Once the nature of these building blocks of profitability is understood, we are able to contruct reserving and investment strategies that increase the ability of an insurer's management to achieve their goals in spite of a volatile, and often hostile, economic environment. The reserving strategy utilizes expected loss payment patterns as well as allowing for unforseen contingencies, such as unanticipated inflation. The investment strategy is also based upon an understanding of the loss payment patterns of insurance liabilities. After providing for a sufficient market value of assets to cover both expected loss payments and a reasonable contingency reserve, the investment strategy allows the portfolio manager to strive for larger returns by investing "true" surplus funds in riskier investments.

While this paper describes many useful asset/management concepts, by explaining them in terms of industry aggregates it appears to have ignored the situation of any individual company. The concepts, however, apply with equal force to individual company situations. Once an individual analysis is completed, if an asset/liability mismatch is detected and company management desires to redress the mismatch, there are many rebalancing techniques that are available. Many of these techniques are most easily used in an environment of stable or declining interest rates (from an *immediate* GAAP or Statutory accounting viewpoint). Some techniques have the added advantage of favorable accounting treatment in periods of relatively unstable or high interest rates. However, all techniques in any interest rate environment will add to the *long-range* health of the industry, but only if they are used *before* the disease of high interest rates strikes again.

APPENDIX

GLOSSARY OF INSURANCE AND INVESTMENT TERMS

Adjustment Expense

Expenses of investigating and settling claims, these include allocated claim expense and unallocated claim expense. Allocated expenses are comprised of court costs, fees and expenses of independent adjusters, lawyers, witnesses and other expenses which can be charged to specific claims. Unallocated expenses represent salaries and other overhead which cannot be charged to specific claims.

Annual Statement (Convention Blank)

Annual filing an insurer must make with its state insurance commissioner. The annual statement is used to monitor the solvency of insurance companies and it contains many sections, known as "schedules," which cover such things as financial position (balance sheet and income statement), amount and type of insurance reserves, organizational structure and a listing of financial assets.

Automobile Liability Insurance

Protection for the insured against financial loss because of legal liability for car-related injuries to others or damage to their property.

Automobile Physical Damage Insurance

Coverage for damages or loss of automobile of policyholder, resulting from collision, fire, theft and other perils.

Boiler and Machinery Insurance

Coverage for loss arising out of the operation of pressure, mechanical and electrical equipment. It may cover loss to the boiler and machinery itself, damage to other property, and business interruption losses.

Book Value (Statement Value)

The carrying value of financial assets as prescribed by statutory accounting and the NAIC. Book value is generally historical cost (plus accretion of any discount or minus amortization of any premium) for most fixed income items, including non-defaulted bonds and sinking fund preferred stocks. Common stock is carried at current market value.

Catastrophe

In insurance, a term applied to an incident or series of related incidents involving a loss of more than a million dollars.

Claim (Loss)

A demand for payment of a policy benefit because of the occurrence of an insured event such as death, injury, destruction or damage.

Combined Ratio

A combination of the underwriting expense ratio and the loss and loss expense ratio. A combined ratio under 100 percent indicates an underwriting profit; one over 100 percent generally indicates an underwriting loss.

Commercial Multiple Peril
A package type of insurance for the commercial establishment that includes a wide range of essential coverage.

Comprehensive Personal Liability Insurance
Protection for an insured against loss arising out of legal liability to pay money for damage or injury to others for which the insured is responsible. Does not include automobile liability or business operations.

Discounting
Recognizing the time value of money in determining claim reserves.

Duration
A more exact measure of the "length" of a financial instrument than traditional measures such as "maturity" or "average life." Duration has two distinct interpretations. First, it is the amount of time that must pass before the accumulated effect of a sudden change in interest rates on cash flow reinvestment will exactly offset the price effect of this change on the present value of the remaining cash flows. This interpretation indicates the time horizon over which a rate of return may be guaranteed, or "immunized." The second interpretation of duration indicates the instantaneous change in the market value of a series of cash flows for a given change in interest rates.

Duration Gap
The differential between a target account's actual duration and its target, desired, duration. The target account is the variable that an asset/liability manager wishes to control. Typical target accounts include capital adequacy (either absolute surplus or a ratio of surplus to total assets) and total return on equity.

Earned Premium
Pro-rata share of premium for coverage that falls within the current accounting period.

Expense Ratio
Ratio of underwriting expenses to total premiums written.

Fire Insurance
Coverage for losses caused by fire and lightning, plus resultant damage caused by smoke and water.

GAAP
Generally Accepted Accounting Principles. GAAP for P&C insurance companies varies from statutory accounting practice by allowing the recognition of some expenses to be deferred to later periods.

General Liability
A form of coverage that pertains, for the most part, to claims arising out of the insured's liability for injuries or damage caused by ownership of property, manufacturing operations, contracting operations, sale or distribution of products, and the operation of elevators and the like, as well as professional services.

IBNR

Incurred But Not Reported. Refers to insured losses that have already occurred but have not yet been reported to the insurer.

Loss Expense

See Adjustment Expense.

Loss and Loss Expense Ratio

Ratio of losses and loss expenses incurred to premiums earned.

Loss Reserve (Liability for Unpaid Claims)

The amount needed to provide for the estimated ultimate cost of claims relating to insured events that have occurred on or before a particular date (ordinarily, the balance sheet date). The reserve includes the amount of money required for payments on claims that have already been reported to the insurer, as well as the amount required for IBNR claims.

Malpractice Insurance

Coverage for a professional practitioner, such as a doctor or a lawyer, against liability claims resulting from alleged malpractice in the performance of the insured's services.

MSVR

Mandatory Securities Valuation Reserve. The MSVR is a reserve category that appears on a life insurance company's NAIC Annual Statement. The purpose of the reserve is to allow for the contingency of asset default or capital loss on disposal. The amount of the reserve is determined by a formula which is dependent upon the type of asset and the asset's rating.

NAIC

National Association of Insurance Commissioners. An organization of state insurance officials which attempts to provide national guidelines and suggestions for the administrators of the various state insurance departments. This body has no legislative nor direct regulatory authority, but does play a major role in the development of the laws governing the industry.

Peril

The cause of a possible loss. Perils include fire, windstorm, collision, hail, etc., and are sometimes also referred to as "hazards."

Put (Protective Put)

A put is the right to sell a security at a predetermined price (strike). Puts, like calls or any other option, can be either sold (written) or purchased. A Protective Put describes a package of an underlying security and a put that has been purchased to protect against any market deterioration.

Schedules O and P

The Annual Statement filed by an insurance company is divided into schedules. Schedules O and P pertain to the loss development patterns for historical reserves and the adequacy of reserves for insurance losses.

Statutory Accounting Practice

Accounting practices prescribed or permitted by insurance regulatory authorities. Statutory accounting generally does not allow expenses to be amortized, rather it requires them to be expensed in the year they can be determined. Both statutory and GAAP accounting for P&C companies allow many assets to be carried on a historical cost basis, as opposed to a market value basis (see "Book Value").

Statutory Policyholders' Surplus

The difference between assets and liabilities (that is, "net worth") as defined by statutory accounting practice.

Stock to Surplus Ratio

The ratio of the market value of an insurer's holdings of common stock to the value of its statutory policyholders' surplus. The ratio indicates the vulnerability of surplus to a decline in stock market values.

Strict Liability

Also known as Absolute Liability, the principle of strict liability causes an insurer to be liable for damage caused by certain conditions that are apt to get out of hand or to cause damage, regardless of whether the insurer is at fault.

Unearned Premium

Pro-rata share of premium for coverage that falls within future accounting periods.

REFERENCES

1. Babbel, David F. "Real Immunization with Indexed Bonds." *Financial Analysts Journal*, November/December 1984.
2. Bierwag, G.O.; Kaufman, G.G.; and Toevs, A.L. "Duration: Its Development and Use in Bond Portfolio Management." *Financial Analysts Journal*, July/August 1983.
3. Biger, Nahum and Kahane, Yehuda. "Purchasing Power Risk and the Performance of Non Life Insurance Companies." *Journal of Risk and Insurance*, June 1976.
4. Cummins, J. David and Nye, David J. "Portfolio Optimization Models for Property-Liability Insurance Companies: An Analysis and Some Extensions." *Management Science*, Vol. 27, No. 4, April 1981.
5. Doherty, Neil A. "A Portfolio Theory of Insurance Capacity." *Journal of Risk and Insurance*, September 1982.
6. Ferguson, Ronald E. "Duration." *Proceedings of the Casualty Actuarial Society*, Vol. 70, 1983.
7. Fredman, Albert J. and Moore, Randal J. "The Money Market/Options Portfolio: A Prudent Strategy." *Best's Review—Property/Casualty Edition*, September 1982.
8. Haney, William C. "A Risk Controlled Approach to Managing Corporate Cash Pools." Morgan Stanley, forthcoming.
9. Jaffe, Alfred I. "Points & Viewpoints: Cash Flow—A Two-Way Street." *National Underwriter—Property/Casualty Edition*, May 23, 1980.
10. Jaffe, Alfred I. "Points & Viewpoints: The Cash Flow of the Long Tail." *National Underwriter—Property/Casualty Edition*, November 28, 1980.
11. Lobdell, Leighton M. "Investment Planning at Continental." *Best's Review—Property/Casualty Edition*, September 1979.
12. Lovestead, Scott H. "Management of Financial Exposure for Multi-Line Insurers." Master's Thesis, Massachusetts Institute of Technology, May 1984.
13. Moore, Brian C. "The Effects of Chronic Inflation On a Property/Casualty Insurance Company." *Best's Review—Property/Casualty Edition*, March 1980.
14. Morais, Richard. "Discounting the Downtrodden." *Forbes*, February 25, 1985.
15. Platt, Robert B. and Latainer, Gary D. "Replicating Option Strategies for Portfolio Risk Control." Morgan Stanley, January 1984.
16. Rubenstein, M. and Leland, H. "Replicating Options with Positions in Stock and Cash." *Financial Analysts Journal*, July/August 1981.
17. Tilley, James A. and Jacob, David P. "Asset/Liability Management for Insurance Companies—Risk Control Techniques." Morgan Stanley, November 1983.
18. Toevs, Alden L. "Uses of Duration Analysis for the Control of Interest Rate Risk." Morgan Stanley, January 1984.
19. Toevs, Alden L. and Haney, William C. "Measuring and Managing Interest Rate Risk: A Guide to Asset/Liability Models Used in Banks and Thrifts." Morgan Stanley, October 1984.

20. "1983 Casualty Loss Reserves." *Best's Insurance Management Reports—On-Line Reports—Property/Casualty,* November 12, 1984.
21. "1983 Casualty Loss Reserves." *Best's Insurance Management Reports—On-Line Reports—Property/Casualty,* December 24, 1984.
22. "Recognition of the Time Value of Money in Loss Reserves." American Academy of Actuaries, Statement 1980-12, Paper #1, April 16, 1980.
23. "Discounting of Loss Reserves." American Academy of Actuaries, Statement 1980-12, paper #2, April 16, 1980.
24. "Discounting Loss Reserves—GAAP vs. Statutory." American Academy of Actuaries, Statement 1980-12, Paper #3, April 16, 1980.
25. "Discounting Claims of Insurance Enterprises." American Institute of Certified Public Accountants, Draft Issues Paper, March 1, 1983.
26. "Appropriateness of a Mandatory Securities Valuation Reserve Concept for Property and Casualty Insurers." Tillinghast, Nelson & Warren Inc., prepared for the NAIC, February 1978.
27. "Appropriateness of a Mandatory Securities Valuation Reserve Concept for Property and Casualty Insurers—Phase II: Procedure." Tillinghast, Nelson & Warren Inc., prepared for the NAIC, March 1979.
28. "Final Report of the Joint Committee on the Role of the Valuation Actuary in the United States." American Academy of Actuaries; Society of Actuaries, February 1985.
29. "Congress Should Consider Changing Federal Income Taxation of the Property/Casualty Insurance Industry." United States General Accounting Office, Comptroller General's Report to the Chairman/Committee on Finance/United States Senate, Draft.
30. "1983 Casualty Loss Reserve Seminar." American Academy of Actuaries/Casualty Actuarial Society, Transcript, September 12-13, 1983.
31. *Best's Aggregates & Averages—Property/Casualty,* 1973 through 1984 editions, A.M. Best Company, Oldwick, New Jersey.
32. Annual Reports: The Progressive Corporation 1981, 1982 and 1983.
    Berkshire Hathaway Corp. 1980.

CAS Exam 9 Study Kit

"Rethinking Risk Management"

by R.M. Stulz

RETHINKING RISK MANAGEMENT

by René M. Stulz,
*The Ohio State University**

This article explores an apparent conflict between the theory and current practice of corporate risk management. Academic theory suggests that some companies facing large exposures to interest rates, exchange rates, or commodity prices can increase their market values by using derivative securities to reduce their exposures. The primary emphasis of the theory is on the role of derivatives in reducing the variability of corporate cash flows and, in so doing, reducing various costs associated with financial distress.

The actual corporate use of derivatives, however, does not seem to correspond closely to the theory. For one thing, large companies make far greater use of derivatives than small firms, even though small firms have more volatile cash flows, more restricted access to capital, and thus presumably more reason to buy protection against financial trouble. Perhaps more puzzling, however, is that many companies appear to be using risk management to pursue goals other than reducing variance.

Does this mean that the prevailing academic theory of risk management is wrong, and that "variance-minimization" is not a useful goal for companies using derivatives? Or, is the current corporate practice of risk management misguided and in urgent need of reform? In this paper, I answer "no" to both questions while at the same time suggesting there may be room for improvement in the theory as well as the practice of risk management.

The paper begins by reviewing some evidence that has accumulated about the current practice of corporate risk management. Part of this evidence takes the form of recent "anecdotes," or cases, involving large derivatives losses. Most of the evidence, however, consists of corporate responses to surveys. What the stories suggest, and the surveys seem to confirm, is the popularity of a practice known as "selective" as opposed to "full-cover" hedging. That is, while few companies regularly use derivatives to take a "naked" speculative position on FX rates or commodity prices, most corporate derivatives users appear to allow their views of future interest rates, exchange rates, and commodity prices to influence their hedge ratios.

Such a practice seems inconsistent with modern risk management theory, or at least the theory that has been presented thus far. But there is a plausible defense of selective hedging—one that would justify the practice without violating the efficient markets tenet at the center of modern financial theory. In this paper, I attempt to explain more of the corporate behavior we observe by pushing the theory of risk management beyond the variance-minimization model that prevails in most academic circles. Some companies, I argue below, may have a comparative advantage in bearing certain financial risks (while other companies mistakenly think and act as if they do). I accordingly propose a somewhat different goal for corporate risk management—namely, the *elimination of costly lower-tail outcomes*—that is designed to reduce the expected costs of financial trouble while preserving a company's ability to exploit any comparative advantage in risk-bearing it may have. (In the jargon of finance specialists, the fundamental aim of corporate risk management can be viewed as the purchase of "well-out-of-the-money put options" that eliminate the downside while preserving as much of the upside as can be justified by the principle of comparative advantage.)

*I am grateful for extensive editorial assistance from Don Chew, and for comments by Steve Figlewski, Andrew Karolyi, Robert Whaley, and participants at a seminar at McKinsey, at the Annual Meeting of the International Association of Financial Engineers, and at the French Finance Association.

Such a modified theory of risk management implies that some companies should hedge all financial risks, other firms should worry about only certain kinds of risks, and still others should not worry about risks at all. But, as I also argue below, when making decisions whether or not to hedge, management should keep in mind that risk management can be used to change both a company's capital structure and its ownership structure. By reducing the probability of financial trouble, risk management has the potential both to increase debt capacity and to facilitate larger equity stakes for management.

This paper also argues that common measures of risk such as variance and Value at Risk (VaR) are not useful for most risk management applications by non-financial companies, nor are they consistent with the objective of risk management presented here. In place of both VaR and the variance of cash flows, I suggest a method for measuring corporate exposures that, besides having a foundation in modern finance theory, should be relatively easy to use.

I conclude with a discussion of the internal "management" of risk management. If corporate risk management is focused not on minimizing variance, but rather on eliminating downside risk while extending the corporate quest for comparative advantage into financial markets, then much more attention must be devoted to the evaluation and control of corporate risk-management activities. The closing section of the paper offers some suggestions for evaluating the performance of risk managers whose "view-taking" is an accepted part of the firm's risk management strategy.

RISK MANAGEMENT IN PRACTICE

In one of their series of papers on Metallgesellschaft, Chris Culp and Merton Miller make an observation that may seem startling to students of modern finance: "We need hardly remind readers that most value-maximizing firms do not hedge."[1] But is this true? And, if so, how would we know?

Culp and Miller refer to survey evidence—in particular, to a Wharton-Chase study that sent questionnaires to 1,999 companies inquiring about their risk management practices.[2] Of the 530 firms that responded to the survey, only about a third answered "yes" when asked if they ever used futures, forwards, options, or swaps. One clear finding that emerges from this survey is that large companies make greater use of derivatives than smaller firms. Whereas 65% of companies with a market value greater than $250 million reported using derivatives, only 13% of the firms with market values of $50 million or less claimed to use them.

What are the derivatives used to accomplish? The only uses reported by more than half of the corporate users are to hedge contractual commitments and to hedge anticipated transactions expected to take place within 12 months. About two thirds of the companies responded that they never use derivatives to reduce funding costs (or earn "treasury profits") by arbitraging the markets or by taking a view. Roughly the same proportion of firms also said they never use derivatives to hedge their balance sheets, their foreign dividends, or their economic or competitive exposures.

The Wharton-Chase study was updated in 1995, and its results were published in 1996 as the Wharton-CIBC Wood Gundy study. The results of the 1995 survey confirm those of its predecessor, but with one striking new finding: Over a third of all derivative users said they sometimes "actively take positions" that reflect their market views of interest rate and exchange rates.

This finding was anticipated in a survey of Fortune 500 companies conducted by Walter Dolde in 1992, and published in this journal in the following year.[3] Of the 244 companies that responded to Dolde's survey, 85% reported having used swaps, forwards, futures, or options. As in the Wharton surveys, larger companies reported greater use of derivatives than smaller firms. And, as Dolde notes, such a finding confirms the experience of risk management practitioners that the corporate use of derivatives requires a considerable upfront investment in personnel, training, and computer hardware and software—an investment that could discourage small firms.

1. Christopher Culp and Merton Miller, "Hedging in the Theory of Corporate Finance: A Reply to Our Critics," *Journal of Applied Corporate Finance* Vol. 8 No. 1 (Spring 1995), p. 122. For the central idea of this paper, I am indebted to Culp and Miller's discussion of Holbrook Working's "carrying-charge" theory of commodity hedging. It is essentially Working's notion—and Culp and Miller's elaboration of it—that I attempt in this paper to generalize into a broader theory of risk management based on comparative advantage in risk-bearing.

2. The Wharton School and The Chase Manhattan Bank, N.A., "Survey of Derivative Usage Among U.S. Non-Financial Firms" (February 1994).

3. Walter Dolde, "The Trajectory of Corporate Financial Risk Management," *Journal of Applied Corporate Finance*, Vol. 6 No. 4 (Fall 1993), 33-41.

But, as we observed earlier, there are also reasons why the demand for risk management products should actually be greater for small firms than for large—notably the greater probability of default caused by unhedged exposures and the greater concentration of equity ownership in smaller companies. And Dolde's survey provides an interesting piece of evidence in support of this argument. When companies were asked to estimate what percentages of their exposures they chose to hedge, many respondents said that it depended on whether they had a view of future market movements. *Almost 90% of the derivatives users in Dolde's survey said they sometimes took a view.* And, when the companies employed such views in their hedging decisions, the smaller companies reported hedging significantly greater percentages of their FX and interest rate exposures than the larger companies.

Put another way, the larger companies were more inclined to "self-insure" their FX or interest rate risks. For example, if they expected FX rates to move in a way that would increase firm value, they might hedge only 10% to 20% (or maybe none) of their currency exposure. But if they expected rates to move in a way that would reduce value, they might hedge 100% of the exposure.

Like the Wharton surveys, the Dolde survey also found that the focus of risk management was mostly on transaction exposures and near-term exposures. Nevertheless, Dolde also reported "a distinct evolutionary pattern" in which many firms "progress from targeting individual transactions to more systematic measures of ongoing competitive exposures."[4]

The bottom line from the surveys, then, is that corporations do not systematically hedge their exposures, the extent to which they hedge depends on their views of future price movements, the focus of hedging is primarily on near-term transactions, and the use of derivatives is greater for large firms than small firms. Many of the widely-reported derivative problems of recent years are fully consistent with this survey evidence, and closer inspection of such cases provides additional insight into common risk management practices. We briefly recount two cases in which companies lost large amounts of money as a result of risk management programs.

Metallgesellschaft

Although the case of Metallgesellschaft continues to be surrounded by controversy, there is general agreement about the facts of the case. By the end of 1993, MGRM, the U.S. oil marketing subsidiary of Metallgesellschaft, contracted to sell 154 million barrels of oil through fixed-price contracts ranging over a period of ten years. These fixed-price contracts created a huge exposure to oil price increases that MGRM decided to hedge. However, it did not do so in a straightforward way. Rather than hedging its future outflows with offsetting positions of matching maturities, MGRM chose to take "stacked" positions in short-term contracts, both futures and swaps, and then roll the entire "stack" forward as the contracts expired.

MGRM's choice of short-term contracts can be explained in part by the lack of longer-term hedging vehicles. For example, liquid markets for oil futures do not go out much beyond 12 months. But it also appears that MGRM took a far larger position in oil futures than would have been consistent with a variance-minimizing strategy. For example, one study estimated that the minimum-variance hedge position for MGRM would have required the forward purchase of only 86 million barrels of oil, or about 55% of the 154 million barrels in short-maturity contracts that MGRM actually entered into.[5]

Does this mean that MGRM really took a position that was long some 58 million barrels of oil? Not necessarily. As Culp and Miller demonstrate, had MGRM adhered to its professed strategy and been able to obtain funding for whatever futures losses it incurred over the entire 10-year period, its position would have been largely hedged.[6]

But even if MGRM's net exposure to oil prices was effectively hedged over the long haul, it is also clear that MGRM's traders had not designed their hedge with the aim of minimizing the variance of their net position in oil during the life of the contracts. The traders presumably took the position they did because they thought they could benefit from their specialized information about supply and demand—and, more specifically, from a persistent feature of oil futures known as "backwardation," or the long-run

4. Dolde, p. 39.

5. Mello, A., and J.E. Parsons, "Maturity Structure of a Hedge Matters: Lessons from the Metallgesellschaft Debacle," *Journal of Applied Corporate Finance*, Vol. 8 No. 1 (Spring 1995), 106-120.

6. More precisely, Culp and Miller's analysis shows that, ignoring any complications arising from basis risk and the daily mark-to-market requirement for futures, over the 10-year period each rolled-over futures contract would have eventually corresponded to an equivalent quantity of oil delivered to customers.

The lesson of market efficiency for corporate risk managers is that the attempt to earn higher returns in most financial markets generally means bearing large (and unfamiliar) risks.

tendency of spot prices to be higher than futures prices. So, although MGRM was effectively hedged against changes in spot oil prices, it nevertheless had what amounted to a long position in "the basis." Most of this long position in the basis represented a bet that the convenience yields on crude oil—that is, the premiums of near-term futures over long-dated futures—would remain positive as they had over most of the past decade.

When spot prices fell dramatically in 1993, MGRM lost on its futures positions and gained on its cash positions—that is, on the present value of its delivery contracts. But because the futures positions were marked to market while the delivery contracts were not, MGRM's financial statements showed large losses. Compounding this problem of large "paper losses," the backwardation of oil prices also disappeared, thus adding real losses to the paper ones. And, in response to the reports of mounting losses, MG's management chose to liquidate the hedge. This action, as Culp and Miller point out, had the unfortunate consequence of "turning paper losses into realized losses" and "leaving MGRM exposed to rising prices on its remaining fixed-price contracts."[7]

Daimler-Benz

In 1995, Daimler-Benz reported first-half losses of DM1.56 billion, the largest in the company's 109-year history. In its public statements, management attributed the losses to exchange rate losses due to the weakening dollar. One subsidiary of Daimler-Benz, Daimler-Benz Aerospace, had an order book of DM20 billion, of which 80% was fixed in dollars. Because the dollar fell by 14% during this period, Daimler-Benz had to take a provision for losses of DM1.2 billion to cover future losses.

Why did Daimler-Benz fail to hedge its expected dollar receivables? The company said that it chose not to hedge because the forecasts it received were too disperse, ranging as they did from DM1.2 to DM1.7 per dollar. Analysts, however, attributed Daimler-Benz's decision to remain unhedged to its view that the dollar would stay above DM1.55.[8]

These two brief case studies reinforce the conclusion drawn from the survey evidence. In both of these cases, management's view of future price movements was an important determinant of how (or whether) risk was managed. Risk management did not mean minimizing risk by putting on a minimum-variance hedge. Rather, it meant choosing to bear certain risks based on a number of different considerations, including the belief that a particular position would allow the firm to earn abnormal returns.

Is such a practice consistent with the modern theory of risk management? To answer that question, we first need to review the theory.

THE PERSPECTIVE OF MODERN FINANCE

The two pillars of modern finance theory are the concepts of efficient markets and diversification. Stated as briefly as possible, market efficiency means that markets don't leave money on the table. Information that is freely accessible is incorporated in prices with sufficient speed and accuracy that one cannot profit by trading on it.

Despite the spread of the doctrine of efficient markets, the world remains full of corporate executives who are convinced of their own ability to predict future interest rates, exchange rates, and commodity prices. As evidence of the strength and breadth of this conviction, many companies during the late '80s and early '90s set up their corporate treasuries as "profit centers" in their own right—a practice that, if the survey evidence can be trusted, has been largely abandoned in recent years by most industrial firms. And the practice has been abandoned with good reason: Behind most large derivative losses—in cases ranging from Orange County and Baring Brothers to Procter & Gamble and BancOne—there appear to have been more or less conscious decisions to bear significant exposures to market risks with the hope of earning abnormal returns.

The lesson of market efficiency for corporate risk managers is that the attempt to earn higher returns in most financial markets generally means bearing large (and unfamiliar) risks. In highly liquid markets such as those for interest rate and FX futures—and in the case of heavily traded commodities like oil and gold as well—industrial companies are unlikely to have a comparative advantage in bearing these risks. And so, for most industrial corporations, setting up the corporate treasury to

7. Culp and Miller, Vol. 7 No. 4 (Winter 1995), p. 63.
8. See *Risk Magazine*, October 1995, p. 11.

trade derivatives for profit is a value-destroying proposition. (As I will also argue later, however, market efficiency does not rule out the possibility that management's information may be better than the market's in special cases.)

But if the concept of market efficiency should discourage corporations from *creating* corporate exposures to financial market risks, the companion concept of diversification should also discourage some companies from *hedging* financial exposures incurred through their normal business operations. To explain why, however, requires a brief digression on the corporate cost of capital.

Finance theory says that the stock market, in setting the values of companies, effectively assigns minimum required rates of return on capital that vary directly with the companies' levels of risk. In general, the greater a company's risk, the higher the rate of return it must earn to produce superior returns for its shareholders. But a company's required rate of return, also known as its cost of capital, is said to depend only on its non-diversifiable (or "systematic") risk, not on its total risk. In slightly different words, a company's cost of capital depends on the strength of the firm's tendency to move with the broad market (in statistical terms, its "covariance") rather than its overall volatility (or "variance").

In general, most of a company's interest rate, currency, and commodity price exposures will not increase the risk of a well-diversified portfolio. Thus, most corporate financial exposures represent "non-systematic" or "diversifiable" risks that shareholders can eliminate by holding diversified portfolios. And because shareholders have such an inexpensive risk-management tool at their disposal, companies that reduce their earnings volatility by managing their financial risks will not be rewarded by investors with lower required rates of return (or, alternatively, with higher P/E ratios for given levels of cash flow or earnings). As one example, investors with portfolios that include stocks of oil companies are not likely to place higher multiples on the earnings of petrochemical firms just because the latter smooth their earnings by hedging against oil price increases.

For this reason, having the corporation devote resources to reducing FX or commodity price risks makes sense only if the cash flow variability arising from such risks has the potential to impose "real" costs on the corporation. The academic finance literature has identified three major costs associated with higher variability: (1) higher expected bankruptcy costs (and, more generally, costs of financial distress); (2) higher expected payments to corporate "stakeholders" (including higher rates of return required by owners of closely-held firms); and (3) higher expected tax payments. The potential gains from risk management come from its ability to reduce each of these three costs—and I review each in turn below.[9]

Risk Management Can Reduce Bankruptcy Costs

Although well-diversified shareholders may not be concerned about the cash flow variability caused by swings in FX rates or commodity prices, they will become concerned if such variability materially raises the probability of financial distress. In the extreme case, a company with significant amounts of debt could experience a sharp downturn in operating cash flow—caused in part by an unhedged exposure—and be forced to file for bankruptcy.

What are the costs of bankruptcy? Most obvious are the payments to lawyers and court costs. But, in addition to these "direct" costs of administration and reorganization, there are some potentially larger "indirect" costs. Companies that wind up in Chapter 11 face considerable interference from the bankruptcy court with their investment and operating decisions. And such interference has the potential to cause significant reductions in the ongoing operating value of the firm.

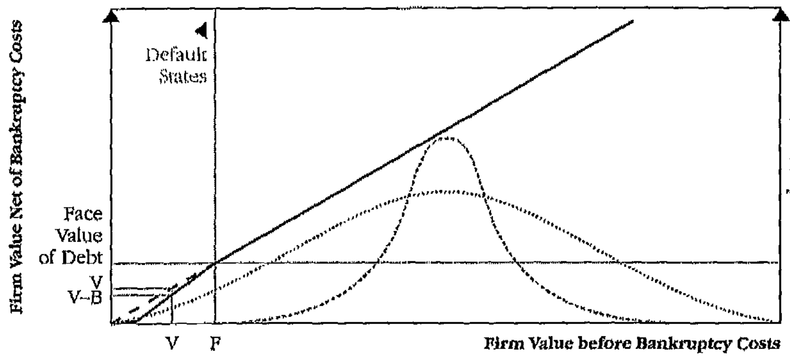
If a company's shareholders view bankruptcy as a real possibility—and to the extent the process of reorganization itself is expected to reduce the firm's operating value—the expected present value of these costs will be reflected in a company's *current* market value. A risk management program that costlessly eliminates the risk of bankruptcy effectively reduces these costs to zero and, in so doing, increases the value of the firm.

The effects of risk management on bankruptcy costs and firm value are illustrated in Figure 1. In the case shown in the figure, hedging is assumed to

9. For a discussion of the benefits of corporate hedging, see Clifford Smith and René Stulz, "The Determinants of Firms' Hedging Policies," *Journal of Financial and Quantitative Analysis* 20 (1985), pp. 391-405.

Because shareholders have such an inexpensive risk-management tool, companies that reduce their earnings volatility by managing financial risks will not be rewarded with a lower "cost of capital." But if shareholders are not concerned about the cash flow variability caused by swings in FX rates or commodity prices, they will become concerned if such variability materially raises the probability of financial distress.

FIGURE 1
DEBT, EQUITY, AND FIRM VALUE WITH BANKRUPTCY COSTS



reduce the variability of cash flow and firm value to the degree that default is no longer possible. By eliminating the possibility of bankruptcy, risk management increases the value of the firm's equity by an amount roughly equal to Bc (bankruptcy costs) multiplied by the probability of bankruptcy if the firm remains unhedged (pBU). For example, let's assume the market value of the firm's equity is $100 million, bankruptcy costs are expected to run $25 million (or 25% of current firm value), and the probability of bankruptcy in the absence of hedging is 10%. In this case, risk management can be seen as increasing the current value of the firm's equity by $2.5 million (10% x $25 million), or 2.5%. (Keep in mind that this is the contribution of risk management to firm value *when the company is healthy*; in the event that cash flow and value should decline sharply from current levels, the value added by risk management increases in absolute dollars, and even more on a percentage-of-value basis.)

This argument extends to distress costs in general. For instance, as a company becomes weaker financially, it becomes more difficult for it to raise funds. At some point, the cost of outside funding—if available at all—may become so great that management chooses to pass up profitable investments. This "underinvestment problem" experienced by companies when facing the prospect of default (or, in some cases, just a downturn in earnings[10]) represents an important cost of financial distress. And, to the extent that risk management succeeds in reducing the perceived *probability* of financial distress and the costs associated with underinvestment, it will increase the current market value of the firm.

Risk Management Can Reduce Payments to "Stakeholders" (and Required Returns to Owners of Closely Held Firms)[11]

Although the shareholders of large public companies can often manage most financial risks more efficiently than the companies themselves, the case may be different for the owners—or owner-managers—of private or closely-held companies. Because such owners tend to have a large proportion of their wealth tied up in the firm, their required rates of return are likely to reflect all important sources of risk, those that can be "diversified away" by outside investors as well as those that cannot. In such circumstances, hedging financial exposures can be thought of as adding value by reducing the owners' risks and hence their required rates of return on investment.

And it's not just the owners of closely held companies that value the protection from risk management. In public companies with dispersed ownership, non-investor groups such as managers, employees, customers, and suppliers with a large stake in the success of the firm typically cannot diversify away large financial exposures. If there is a chance that their "firm-specific" investments could be lost because of financial distress, they are likely to require added compensation for the greater risk.

10. This argument is made by Kenneth Froot, David Scharfstein, and Jeremy Stein in "Risk Management: Coordinating Corporate Investment and Financing Policies," *Journal of Finance* 48, (1993), 1629-1658.

11. The discussion in this section and the next draws heavily on Smith and Stulz (1985), cited in footnote 9.

Employees will demand higher wages (or reduce their loyalty or perhaps their work effort) at a company where the probability of layoff is greater. Managers with alternative opportunities will demand higher salaries (or maybe an equity stake in the company) to run firms where the risks of insolvency and financial embarrassment are significant. Suppliers will be more reluctant to enter into long-term contracts, and trade creditors will charge more and be less flexible, with companies whose prospects are more uncertain. And customers concerned about the company's ability to fulfil warranty obligations or service their products in the future may be reluctant to buy those products.

To the extent risk management can protect the investments of each of these corporate stakeholders, the company can improve the terms on which it contracts with them and so increase firm value. And, as I discuss later in more detail, hedging can also facilitate larger equity stakes for managers of public companies by limiting "uncontrollables" and thus the "scope" of their bets.

Risk Management Can Reduce Taxes

The potential tax benefits of risk management derive from the interaction of risk management's ability to reduce the volatility of reported income and the progressivity (or, more precisely, the "convexity") of most of the world's tax codes. In the U.S., as in most countries, a company's effective tax rate rises along with increases in pre-tax income. Increasing marginal tax rates, limits on the use of tax-loss carry forwards, and the alternative minimum tax all work together to impose higher effective rates of taxation on higher levels of reported income and to provide lower percentage tax rebates for ever larger losses.

Because of the convexity of the tax code, there are benefits to "managing" taxable income so that as much of it as possible falls within an optimal range—that is, neither too high nor too low. By reducing fluctuations in taxable income, risk management can lead to lower tax payments by ensuring that, over a complete business cycle, the largest possible proportion of corporate income falls within this optimal range of tax rates.

RISK MANAGEMENT AND COMPARATIVE ADVANTAGE IN RISK-TAKING

Up to this point, we have seen that companies should not expect to make money consistently by taking financial positions based on information that is publicly available. But what about information that is not publicly available? After all, many companies in the course of their normal operating activities acquire specialized information about certain financial markets. Could not such information give them a comparative advantage over their shareholders in taking some types of risks?

Let's look at a hypothetical example. Consider company X that produces consumer durables using large amounts of copper as a major input. In the process of ensuring that it has the appropriate amount of copper on hand, it gathers useful information about the copper market. It knows its own demand for copper, of course, but it also learns a lot about the supply. In such a case, the firm will almost certainly allow that specialized information to play some role in its risk management strategy.

For example, let's assume that company X's management has determined that, when it has no view about future copper prices, it will hedge 50% of the next year's expected copper purchases to protect itself against the possibility of financial distress. But, now let's say that the firm's purchasing agents persuade top management that the price of copper is far more likely to rise than fall in the coming year. In this case, the firm's risk manager might choose to take a long position in copper futures that would hedge as much as 100% of its anticipated purchases for the year instead of the customary 50%. Conversely, if management becomes convinced that copper prices are likely to drop sharply (with almost no possibility of a major increase), it might choose to hedge as little as 20% of its exposure.[12]

Should the management of company X refrain from exploiting its specialized knowledge in this fashion, and instead adhere to its 50% hedging target? Or should it, in certain circumstances, allow its market view to influence its hedge ratio?

Although there are clearly risks to selective hedging of this kind—in particular, the risk that the

12. For a good example of this kind of selective hedging policy, see the comments by John Van Roden, Chief Financial Officer of Lukens, Inc. in the "Bank of America Roundtable on Corporate Risk Management," *Journal of Applied Corporate Finance*, Vol. 8 No. 3 (Fall 1995). As a stainless steel producer, one of the company's principal inputs is nickel; and Lukens' policy is to allow its view of nickel prices to influence how much of its nickel exposure it hedges. By contrast, although it may have views of interest rates or FX exposures, such views play no role in hedging those exposures.

How can management determine when it should take risks and when it should not? The best approach is to implement a *risk-taking audit*—a comprehensive review of the risks to which the company is exposed, both through its financial instruments and liability structure as well as its normal operations.

firm's information may not in fact be better than the market's—it seems quite plausible that companies could have such informational advantages. Companies that repurchase their own shares based on the belief that their current value fails to reflect the firm's prospects seem to be vindicated more often than not. And though it's true that management may be able to predict the firm's future earnings with more confidence than the price of one of its major inputs, the information companies acquire about certain financial markets may still prove a reasonably reliable source of gain in risk management decisions.

The Importance of Understanding Comparative Advantage

What this example fails to suggest, however, is that the same operating activity in one company may not necessarily provide a comparative advantage in risk-bearing for another firm. As suggested above, the major risk associated with "selective" hedging is that the firm's information may not in fact be better than the market's. For this reason, it is important for management to understand the source of its comparative advantages.

To illustrate this point, take the case of a foreign currency trading operation in a large commercial bank. A foreign currency trading room can make a lot of money from taking positions provided, of course, exchange rates move in the anticipated direction. But, in an efficient market, as we have seen, banks can reliably make money from position-taking of this sort only if they have access to information before most other firms. In the case of FX, this is likely to happen only if the bank's trading operation is very large—large enough so that its deal flow is likely to reflect general shifts in demand for foreign currencies.

Most FX dealers, however, have no comparative advantage in gathering information about changes in the value of foreign currencies. For such firms, management of currency risk means ensuring that their exposures are short-lived. The most reliable way to minimize exposures for most currency traders is to enlarge their customer base. With a sufficient number of large, highly active customers, a trading operation has the following advantage: If one of its traders agrees to buy yen from one customer, the firm can resell them quickly to another customer and pocket the bid-ask spread.

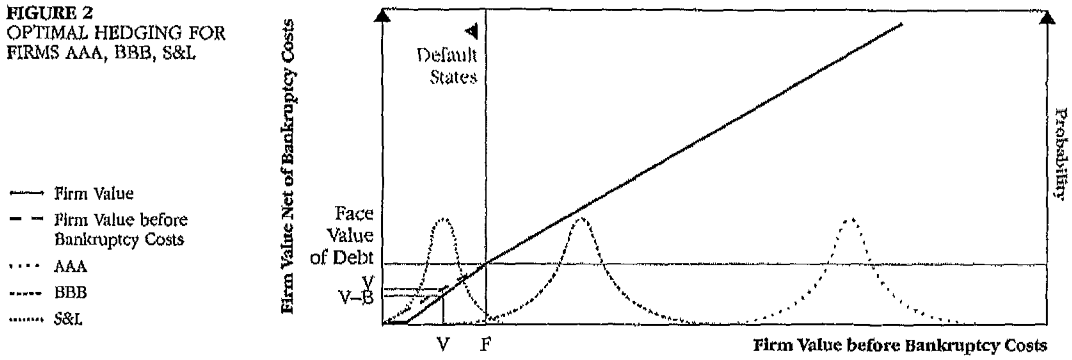
In an article entitled "An Analysis of Trading Profits: How Trading Rooms Really Make Money," Alberic Braas and Charles Bralver present evidence suggesting that most FX trading profits come from market-making, not position-taking.[13] Moreover, as the authors of this article point out, a trading operation that does not understand its comparative advantage in trading currencies is likely not only to fail to generate consistent profit, but to endanger its existing comparative advantage. If the source of the profits of the trading room is really the customer base of the bank, and not the predictive power of its traders, then the bank must invest in maintaining and building its customer base. A trading room that mistakenly believes that the source of its profits is position-taking will take large positions that, on average, will neither make money nor lose money. More troubling, though, is that the resulting variability of its trading income is likely to unsettle its customers and weaken its customer base. Making matters worse, it may choose a compensation system for its traders that rewards profitable position-taking instead of valuable coordination of trading and sales activities. A top management that fails to understand its comparative advantage may waste its time looking for star traders while neglecting the development of marketing strategies and services.

How can management determine when it should take risks and when it should not? The best approach is to implement a *risk-taking audit*. This would involve a comprehensive review of the risks to which the company is exposed, both through its financial instruments and liability structure as well as its normal operations. Such an audit should attempt to answer questions like the following: Which of its major risks has the firm proved capable of "self-insuring" over a complete business cycle? If the firm chooses to hedge "selectively," or leaves exposures completely unhedged, what is the source of the firm's comparative advantage in taking these positions? Which risk management activities have consistently added value without introducing another source of volatility?

Once a firm has decided that it has a comparative advantage in taking certain financial risks, it must then determine the role of risk management

13. See Albéric Braas and Charles Bralver, "An Analysis of Trading Profits: How Most Trading Rooms Really Make Money," *Journal of Applied Corporate Finance*, Vol. 2 No. 4 (Winter 1990).

FIGURE 2
OPTIMAL HEDGING FOR FIRMS AAA, BBB, S&L



in exploiting this advantage. As I argue below, risk management may paradoxically enable the firm to take *more* of these risks than it would in the absence of risk management. To illustrate this point, let's return to our example of company X and assume it has valuable information about the copper market that enables it to earn consistently superior profits trading copper. Even in this situation, such trading profits are by no means a sure thing; there is always the possibility that the firm will experience significant losses. Purchasing far-out-of-the-money calls on copper in such a case could actually serve to increase the firm's ability to take speculative positions in copper. But, as I argue in the next section, a company's ability to withstand large trading losses without endangering its operating activities depends not only on its risk management policy, but also on its capital structure and general financial health.

THE LINK BETWEEN RISK MANAGEMENT, RISK-TAKING, AND CAPITAL STRUCTURE

In discussing earlier the benefits of risk management, I suggested that companies should manage risk in a way that makes financial distress highly unlikely and, in so doing, preserves the financing flexibility necessary to carry out their investment strategies. Given this primary objective for risk management, one would not expect companies with little or no debt financing—and, hence, a low probability of financial trouble—to benefit from hedging.

In this sense, risk management can be viewed as a direct substitute for equity capital. That is, the more the firm hedges its financial exposures, the less equity it requires to support its business. Or, to put it another way, the use of risk management to reduce exposures effectively increases a company's debt capacity.

Moreover, to the extent one views risk management as a substitute for equity capital—or, alternatively, as a technique that allows management to substitute debt for equity—then it pays companies to practice risk management only to the extent that equity capital is more expensive than debt. As this formulation of the issue suggests, a company's decisions to hedge financial risks—or to bear part of such risks through selective hedging—should be made jointly with the corporate capital structure decision.

To illustrate this interdependence between risk management and capital structure, consider the three kinds of companies pictured in Figure 2. At the right-hand side of the figure is company AAA, so named because it has little debt and a very high debt rating. The probability of default is essentially zero; and thus the left or lower tail of AAA's distribution of potential outcomes never reaches the range where low value begins to impose financial distress costs on the firm. Based on the theory of risk management just presented, there is no reason for this company to hedge its financial exposures; the company's shareholders can do the same job more cost-effectively. And, should investment opportunities arise, AAA will likely be able to raise funds on an economic basis, even if its cash flows should decline temporarily.

Should such a company take bets on financial markets? The answer could be yes, provided management has specialized information that would give it a comparative advantage in a certain market. In

The question of what is the right corporate risk management decision for a company begs the question of not only its optimal capital structure, but optimal *ownership* structure as well.

AAA's case, a bet that turns out badly will not affect the company's ability to carry out its strategic plan.

But now let's consider the company in the middle of the picture, call it BBB. Like the company shown in Figure 1 earlier, this firm has a lower credit rating, and there is a significant probability that the firm could face distress. What should BBB do? As shown earlier in Figure 1, this firm should probably eliminate the probability of encountering financial distress through risk management. In this case, even if management feels that there are occasional opportunities to profit from market inefficiencies, hedging exposures is likely to be the best policy. In company BBB's case, the cost of having a bet turn sour can be substantial, since this would almost certainly imply default. Consequently, one would not expect the management of such a firm to let its views affect the hedge ratio.

Finally, let's consider a firm that is in distress—and let's call it "S&L." What should it do? Reducing risk once the firm is in distress is not in the interest of shareholders. If the firm stays in distress and eventually defaults, shareholders will end up with near-worthless shares. In these circumstances, a management intent on maximizing shareholder value will not only accept bets that present themselves, but will *seek out* new ones. Such managers will take bets even if they believe markets are efficient because introducing new sources of volatility raises the probability of the "upper-tail" outcomes that are capable of rescuing the firm from financial distress.

Back to the Capital Structure Decision. As we saw in the case of company AAA, firms that have a lot of equity capital can make bets without worrying about whether doing so will bring about financial distress. One would therefore not expect these firms to hedge aggressively, particularly if risk management is costly and shareholders are better off without it.

The major issue that such companies must address, however, is whether they have too much capital—or, too much equity capital. In other words, although risk management may not be useful to them *given their current leverage ratios*, they might be better off using risk management and increasing leverage. Debt financing, of course, has a tax advantage over equity financing. But, in addition to its ability to reduce corporate taxes, increasing leverage also has the potential to strengthen management incentives to improve efficiency and add value. For one thing, the substitution of debt for equity leads managers to pay out excess capital—an action that could be a major source of value added in industries with overcapacity and few promising investment opportunities. Perhaps even more important, however, is that the substitution of debt for equity also allows for greater concentration of equity ownership, including a significant ownership stake for managers.

In sum, the question of what is the right corporate risk management decision for a company begs the question of not only its optimal capital structure, but optimal *ownership* structure as well. As suggested above, hedging could help some companies to increase shareholder value by enabling them to raise leverage—say, by buying back their shares—and increase management's percentage ownership. For other companies, however, leaving exposures unhedged or hedging "selectively" while maintaining more equity may turn out to be the value-maximizing strategy.

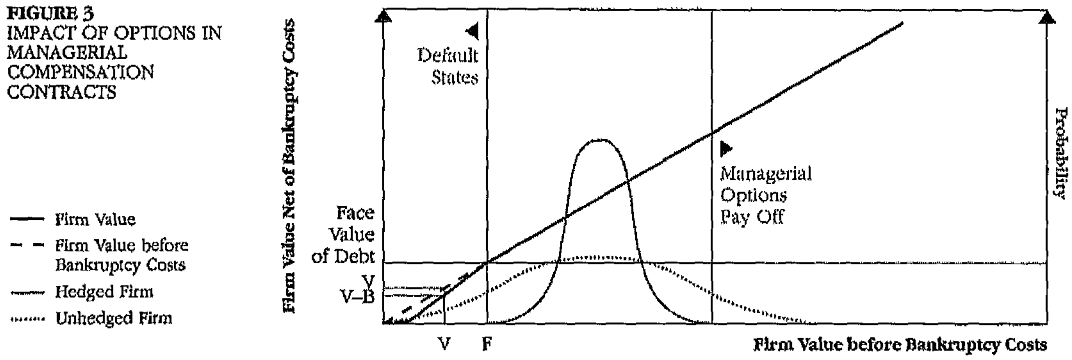
CORPORATE RISK-TAKING AND MANAGEMENT INCENTIVES

Management incentives may have a lot to do with why some firms take bets and others do not. As suggested, some companies that leave exposures unhedged or take bets on financial markets may have a comparative advantage in so doing; and, for those companies, such risk-taking may be a value-increasing strategy. Other companies, however, may choose to take financial risks without having a comparative advantage, particularly if such risk-taking somehow serves the interests of those managers who choose to expose their firms to the risks.

We have little convincing empirical evidence on the extent of risk-taking by companies, whether public or private. But there is one notable exception—a study by Peter Tufano of the hedging behavior of 48 publicly traded North American gold mining companies that was published in the September 1996 issue of the *Journal of Finance*.[14] The gold mining industry is ideal for studying hedging behav-

14. Peter Tufano, "Who Manages Risk? An Empirical Examination of the Risk Management Practices of the Gold Mining Industry," *Journal of Finance* (September, 1996).

FIGURE 3
IMPACT OF OPTIONS IN MANAGERIAL COMPENSATION CONTRACTS



ior in the sense that gold mining companies tend to be single-industry firms with one very large price exposure and a wide range of hedging vehicles, from forward sales, to exchange-traded gold futures and options, to gold swaps and bullion loans.

The purpose of Tufano's study was to examine the ability of various corporate risk management theories to explain any significant pattern of differences in the percentage of their gold price exposures that the companies choose to hedge. Somewhat surprisingly, there was considerable variation in the hedging behavior of these 48 firms. One company, Homestake Mining, chose not only to hedge none of its exposure, but to publicize its policy while condemning what it called "gold price management." At the other extreme were companies like American Barrick that hedged as much as 85% of their anticipated production over the next three years. And whereas about one in six of these firms chose to hedge none of its exposure and sold *all* of its output at spot prices, another one in six firms hedged 40% or more of its gold price exposure.

The bottom line of Tufano's study was that the only important systematic determinant of the 48 corporate hedging decisions was managerial ownership of shares and, more generally, the nature of the managerial compensation contract. In general, the greater management's direct percentage share ownership, the larger the percentage of its gold price exposure a firm hedged. By contrast, little hedging took place in gold mining firms where management owns a small stake. Moreover, managerial compensation contracts that emphasize options or option-like features were also associated with significantly less hedging.

As Tufano acknowledged in his study, this pattern of findings could have been predicted from arguments that Clifford Smith and I presented in a theoretical paper in 1985.[15] Our argument was essentially as follows: As we saw in the case of closely held companies, managers with a significant fraction of their own wealth tied up in their own firms are likely to consider all sources of risk when setting their required rates of return. And this could help explain the tendency of firms with heavy managerial equity ownership to hedge more of their gold price exposures. In such cases, the volatility of gold prices translates fairly directly into volatility of managers' wealth, and manager-owners concerned about such volatility may rationally choose to manage their exposures. (How, or whether, such hedging serves the interests of the companies' outside shareholders is another issue, one that I return to shortly.)

The propensity of managers with lots of stock options but little equity ownership to leave their gold price exposures unhedged is also easy to understand. As shown in Figure 3, the one-sided payoff from stock options effectively rewards management for taking bets and so increasing volatility. In this example, the reduction in volatility from hedging makes management's options worthless (that is, the example assumes these are well out-of-the-money options). But if the firm does not hedge, there is some

15. Clifford Smith and René Stulz, "The Determinants of Firms' Hedging Policies," *Journal of Financial and Quantitative Analysis* 20 (1985), pp. 391-405.

***Given that the firm has chosen to concentrate equity ownership*, hedging may well be a value-adding strategy. If significant equity ownership for managers is expected to strengthen incentives to improve operating performance, hedging can make these incentives even stronger by removing the "noise" introduced by a major performance variable that is beyond management's control.**

probability that a large increase in gold prices will cause the options to pay off.

What if we make the more realistic assumption that the options are *at the money* instead of far out of the money? In this case, options would still have the power to influence hedging behavior because management gains more from increases in firm value than it loses from reductions in firm value. As we saw in the case of the S&L presented earlier, this "asymmetric" payoff structure of options increases management's willingness to take bets.[16]

But if these differences in hedging behavior reflect differences in managerial incentives, what do they tell us about the effect of risk management on shareholder value? Without directly addressing the issue, Tufano implies that neither of the two polar risk management strategies—hedging none of their gold exposure vs. hedging 40% or more—seems designed to increase shareholder value while both appear to serve managers' interests. But can we therefore conclude from this study that neither of these approaches benefits shareholders?

Let's start with the case of the companies that, like Homestake Mining, choose to hedge none of their gold price exposure. As we saw earlier, companies for which financial distress is unlikely have no good reason to hedge (assuming they see no value in changing their current capital structure). At the same time, in a market as heavily traded as gold, management is also not likely to possess a comparative advantage in predicting gold prices. And, lacking either a motive for hedging or superior information about future gold prices, management has no reason to alter the company's natural exposure to gold prices. In further defense of such a policy, one could also argue that such a gold price exposure will have diversification benefits for investors seeking protection against inflation and political risks.

On the other hand, as Smith and I pointed out, because stock options have considerably more upside than downside risk, such incentive packages could result in a misalignment of managers' and shareholders' interests. That is, stock options could be giving managers a one-sided preference for risk-taking that is not fully shared by the companies' stockholders; and, if so, a better policy would be to balance managers' upside potential by giving them a share of the downside risk.

But what about the opposite decision to hedge a significant portion of gold price exposures? Was that likely to have increased shareholder value? As Tufano's study suggests, the managers of the hedging firms tend to hold larger equity stakes. And, as we saw earlier, if such managers have a large fraction of their wealth tied up in their firms, they will demand higher levels of compensation to work in firms with such price exposures. *Given that the firm has chosen to concentrate equity ownership*, hedging may well be a value-adding strategy. That is, if significant equity ownership for managers is expected to add value by strengthening incentives to improve operating performance, the role of hedging is to make these incentives even stronger by removing the "noise" introduced by a major performance variable—the gold price—that is beyond management's control. For this reason, the combination of concentrated ownership, the less "noisy" performance measure produced by hedging, and the possibility of higher financial leverage[17] has the potential to add significant value. As this reasoning suggests, risk management can be used to facilitate an organizational structure that resembles that of an LBO![18]

To put the same thought another way, it is the risk management policy that allows companies with large financial exposures to have significant managerial stock ownership. For, without the hedging policy, a major price exposure would cause the scope of management's bet to be too diffuse, and "uncontrollables" would dilute the desired incentive benefits of more concentrated ownership.

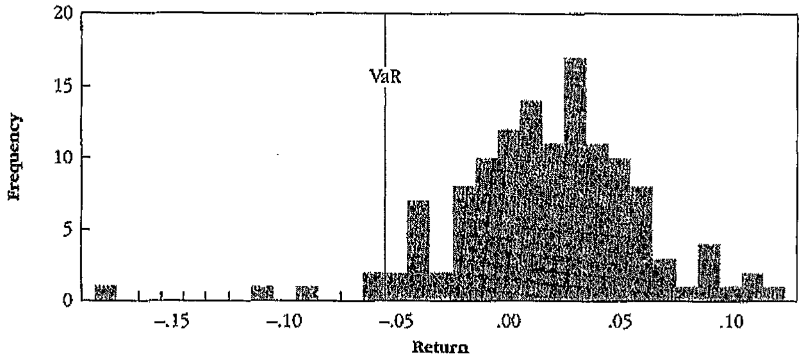
Although Tufano's study is finally incapable of answering the question, "Did risk management add value for shareholders?," the study nevertheless has an important message for corporate policy. It says

16. Additional empirical support for the importance of the relation between the option component of managerial compensation contracts and corporate risk-taking was provided in a recent study of S&Ls that changed their organizational form from mutual ownership to stock ownership. The study finds that those "converted" S&Ls where management has options choose to increase their one-year gaps and, hence, their exposure to interest rates. The study also shows that the greater the percentage of their interest rate exposure an S&L hedges, the larger the credit risk it takes on. The authors of the study interpret this finding to argue, as I do here, that risk management allows firms to increase their exposures to some risks by reducing other risks and thus limiting total firm risk. See C.M. Schrandt and H. Unal, "Coordinated Risk Management: On and Off-balance Sheet Hedging and Thrift Conversion," 1996, unpublished working paper, The Wharton School, University of Pennsylvania, Philadelphia, PA.

17. Although Tufano's study does not find that firms that hedge have systematically higher leverage ratios, it does find that companies that hedge less have higher cash balances.

18. For a discussion of the role of hedging in creating an LBO-like structure, see my study, "Managerial Discretion and Optimal Financing Policies," *Journal of Financial Economics* (1990), pp. 3-26.

FIGURE 4
WORLD MARKET PORTFOLIO RETURN SEPTEMBER 1985-DECEMBER 1995



that, to the extent that risk-taking within the corporation is decentralized, it is important to understand the incentives of those who make the decisions to take or lay off risks.

Organizations have lots of people doing a good job, and so simply doing a good job may not be enough to get promoted. And, if one views corporate promotions as the outcome of "tournaments" (as does one strand of the academic literature), there are tremendous incentives to stand out. One way to stand out is by volunteering to take big risks. In most areas of a corporation, it is generally impossible to take risks where the payoffs are large enough to be noticeable if things go well. But the treasury area may still be an exception. When organized as a profit center, the corporate treasury was certainly a place where an enterprising executive could take such risks and succeed. To the extent such possibilities for risk-taking still exist within some corporate treasuries, top management must be very careful in establishing the appropriate incentives for their risk managers. I return to this subject in the final section of the paper.

MEASURING RISK (OR, IMPROVING ON VaR)

As I mentioned at the outset, the academic literature has focused on volatility reduction as the primary objective of risk management, and on variance as the principal measure of risk. But such a focus on variance, as we have seen, is inconsistent with both most corporate practice and with the theory of risk management presented in this paper. Rather than aiming to reduce variance, most corporate risk management programs appear designed just to avoid "lower-tail outcomes" while preserving upside potential. Indeed, as I suggested earlier, some companies will hedge certain downside risks precisely in order to be able to increase their leverage ratios or to enlarge other financial exposures in ways designed to exploit their comparative advantage in risk-taking.

Many commercial banks and other financial institutions now attempt to quantify the probability of lower-tail outcomes by using a measure known as Value at Risk, or VaR. To illustrate the general principle underlying VaR, let's assume you are an investor who holds a stock portfolio that is fully diversified across all the major world markets. To calculate your Value at Risk, you will need the kind of information that is presented graphically in Figure 4, which is a histogram showing the distribution of monthly returns on the Morgan Stanley Capital International world market portfolio from September 1985 through December 1995.

How risky is that portfolio? One measure is the standard deviation of the portfolio's monthly returns. Over that roughly 10-year period, the average monthly return was 1.23%, with a standard deviation of 4.3%. This tells you that, about two thirds of the time, your actual return would have fallen within a range extending from a loss of 3.1% to a gain of 5.5%.

But what if one of your major concerns is the size of your monthly losses if things turn out badly, and you thus want to know more about the bottom third of the distribution of outcomes? Let's say, for example, that you want to know the maximum extent of your losses in 95 cases out of 100—that is, within a 95% "confidence interval." In that case, you would calculate the VaR evaluated at the 5% level,

The question management would like to be able to answer is this: If we define financial distress as a situation where we cannot raise funds with a rating of BBB, or where our cash flows fall below some target, what is the probability of distress over, say, the next three years? VaR by itself cannot answer this question—nor can traditional measures of volatility.

which turns out to be a loss of 5.9%. This VaR, represented by the vertical line in the middle of Figure 4, is obtained by taking the monthly average return of 1.23% and subtracting from it 1.65 times the standard deviation of 4.3%. And, if you wanted to know the dollar value of your maximum expected losses, you would simply multiply 5.9% times the dollar value of your holdings. That number is your monthly VaR at the 95% confidence level.

Athough the VaR is now used by some industrial firms to evaluate the risks of their derivatives portfolios, the measure was originally designed by J.P. Morgan to help financial institutions monitor the exposures created by their trading activities. In fact, for financial institutions that trade in liquid markets, a *daily* VaR is likely to be even more useful for monitoring trading operations than the monthly VaR illustrated above. Use of a daily VaR would tell an institution that it could expect, in 95 cases out of 100, to lose no more than X% of its value before unwinding its positions.

The special appeal of VaR is its ability to compress the expected distribution of bad outcomes into a single number. But how does one apply such a measure to the corporate risk management we have been discussing? Despite its advantages for certain uses, VaR cannot really be used to execute the risk management goal presented in this paper—namely, the elimination of lower-tail outcomes to avoid financial distress. The fact that there is a 95% probability that a company's loss on a given day, or in a given month, will not exceed a certain amount called VaR is not useful information when management's concern is whether firm value will fall below some critical value *over an extended period of time*. The question management would like to be able to answer is this: If we define financial distress as a situation where we cannot raise funds with a rating of BBB, or where our cash flows or the value of equity fall below some target, what is the probability of distress over, say, the next three years? VaR by itself cannot answer this question—nor can traditional measures of volatility.

It is relatively simple to calculate VaR for a financial institution's portfolio over a horizon of a day or a week. It is much less clear how one would compute the VaR associated with, say, an airline's ongoing operating exposure to oil prices. In evaluating their major risks, most non-financial companies will want to know how much volatility in their cash flows or firm value an exposure can be expected to cause over periods of at least a year, and often considerably longer. Unfortunately, there are at least two major difficulties in extending the VaR over longer time horizons that may not be surmountable.

First, remember that a daily VaR at the 99th percentile is one that is expected to occur on one day out of 100. The relative precision of such a prediction makes it possible to conduct empirical checks of the validity of the model. With the large number of daily observations, one can readily observe the frequency with which the loss is equal or greater than VaR *using reasonably current data*. But, if we attempt to move from a daily to, say, a one-year VaR at the same 99th percentile, it becomes very difficult to calculate such a model, much less subject it to empirical testing. Since an annual VaR at the 99th percentile means that the loss can be expected to take place in only one year in every 100, one presumably requires numerous 100-year periods to establish the validity of such a model.

The second problem in extending the time horizon of VaR is its reliance on the normal distribution. When one is especially concerned about "tail" probabilities—the probabilities of the worst and best outcomes—the assumption made about the statistical distribution of the gains and losses is important. Research on stock prices and on default probabilities across different classes of debt suggests that the tail probabilities are generally larger than implied by the normal distribution. A simple way to understand this is as follows. If stock returns were really normally distributed, as many pricing models assume, market declines in excess of 10% in a day would be extremely rare—say, once in a million years. The fact that such declines happen more often than this is proof that the normal distribution does not describe the probability of lower-tail events correctly.

Although this is not an important failing for most applications in corporate finance, including the valuation of most securities, it can be critical in the context of risk management. For example, if changes in the value of derivatives portfolios or default probabilities have "fatter tails" than those implied by a normal distribution, management could end up significantly understating the probability of distress.

An Alternative to VaR: Using Cash Flow Simulations to Estimate Default Probabilities. Moreover, even if we could calculate a one-year VaR for the value of the firm and be reasonably confident that the distribution was normal, the relevant risk measure for hedging purposes would not be the VaR computed at the one-year horizon. A VaR computed at the one-year horizon at the 99th percentile answers the question: What is the maximum loss in firm value that I can expect in 99 years out of 100? But when a company hedges an exposure, its primary concern is the likelihood of distress *during the year*, which depends on the value of the cumulative loss throughout the year. Thus, it must be concerned about the path of firm value during a period of time rather than the distribution of firm value at the end of the period.

Given this focus on cumulative changes in firm value during a period of time, perhaps the most practical approach to assessing a company's probability of financial distress is to conduct sensitivity analysis on the expected distribution of cash flows. Using Monte Carlo simulation techniques, for example, one could project the company's cash flows over a ten-year horizon in a way that is designed to reflect the combined effect of (and any interactions among) all the firm's major risk exposures on its default probability. The probability of distress over that period would be measured by the fraction of simulated distributions that falls below a certain threshold level of cumulative cash flow. Such a technique could also be used to estimate the expected effect of various hedging strategies on the probability of distress.[19]

One of the advantages of using simulation techniques in this context is their ability to incorporate any special properties (or "non-normalities") of the cash flows. As we saw earlier, the VaR approach assumes that the gains and losses from risky positions are "serially independent," which means that if your firm experiences a loss today, the chance of experiencing another loss tomorrow is unaffected. But this assumption is likely to be wrong when applied to the operating cash flow of a nonfinancial firm: If cash flow is poor today, it is more likely to be poor tomorrow. Simulation has the ability to build this "serial dependence" of cash flows into an analysis of the probability of financial distress.

MANAGING RISK-TAKING

As we have seen, a hedging strategy that focuses on the probability of distress can be consistent with an increase in risk-taking. With such a strategy, the primary goal of risk management is to eliminate lower-tail outcomes. Using risk management in this way, it is possible for a company to increase its volatility while also limiting the probability of a bad outcome that would create financial distress. One example of such a strategy would be to lever up the firm while at the same time buying way out-of-the-money put options that pay off if the firm does poorly. Focusing on lower-tail outcomes is also fully consistent with managing longer-term economic or competitive exposures, as opposed to the near-term transaction exposures that most corporate risk management seems designed to hedge.

But how would the firm decide whether the expected payoff from taking certain financial bets is adequate compensation for not only the risk of losses, but also the expected costs of financial distress? And, once management decides that it is a value-increasing proposition to undertake certain bets, how would the firm evaluate the success of its risk-taking efforts?

To evaluate if the bet is worth taking, let's start by supposing that we are willing to put an explicit cost on the increase in the probability of distress resulting from betting on certain markets. In that case, the trade-off for evaluating a bet for the company becomes fairly simple: The expected profit from the bet must exceed the increase in the probability of distress multiplied by the expected cost of distress.[20] Thus, a bet that has a positive expected value and no effect on the probability of distress is one that the firm should take. But a bet with positive expected profit that significantly increases the probability of financial distress may not appear profitable if the costs of a bad outcome are too large. In such cases, it makes sense for the firm to think

19. For an illustration of the use of Monte Carlo analysis in risk management, see René Stulz and Rohan Williamson, "Identifying and Quantifying Exposures," in Robert Jameson, ed., *Treasury Risk Management* (London, Risk Publications), forthcoming.

20. One possible approach to quantifying the *expected* costs of financial distress involves the concept of American "binary options" and the associated option pricing models. An example of a binary option is one that would pay a fixed amount, say, $10, if the stock price of IBM falls below $40. Unlike standard American put options, which when exercised pay an amount equal to (the strike price of) $40 minus the actual price, the holder of a binary option receives either $10 or nothing, and exercises when the stock price crosses the $40 barrier. Such options can be priced using modified option pricing models.

The connection between binary options and risk management is this: The present value of a binary option is a function of two major variables: the probability that firm value will fall below a certain level (in this case, $40) and the payoff in the event of such a drop in value ($10). By substituting for the $10 payoff its own estimate of how much *additional* value the firm is likely to lose *once its value falls to a certain level and gets into financial trouble*, management can then estimate the expected present value of such costs using a binary option pricing model. This is the number that could be set against the expected profit from the firm's bet in order to evaluate whether to go ahead with the bet.

To the extent that view-taking becomes an accepted part of a company's risk management program, it is important to evaluate managers' bets on a risk-adjusted basis and relative to the market. If managers want to behave like money managers, they should be evaluated like money managers.

about using risk management to reduce the probability of distress. By hedging, management may be able to achieve a reduction in cash flow variability that is large enough that an adverse outcome of the bet will not create financial distress.

Given that management has decided the bet is worth taking, how does it evaluate the outcome of the strategy? Consider first the case of our firm AAA discussed earlier. Recall that this firm is not concerned about lower-tail outcomes and thus has no reason to hedge. When evaluating the outcome of the bet in this case, the appropriate benchmark is the expected gain *adjusted for risk.* It is not enough that the bet ends up earning more than the risk-free rate or even more than the firm's cost of capital. To add value for the company's shareholders, the bet must earn a return that is higher than investors' expected return on other investments of comparable risk.

For example, there is considerable evidence that holding currencies of high-interest rate countries earns returns that, on average, exceed the risk-free rate. This excess return most likely represents "normal" compensation for bearing some kind of risk—say, the higher inflation and interest rate volatility associated with high-interest-rate countries. And because such a strategy is thus *expected* to earn excess returns, it would not make sense to reward a corporate treasury for earning excess returns in this way. The treasury takes risks when it pursues that strategy, and the firm's shareholders expect to be compensated for these risks. Thus, it is only the amount by which the treasury exceeds the expected return—or the "abnormal return"—that represents *economic profit* for the corporation.

So, the abnormal or excess return should be the measure for evaluating bets by company AAA. But now let's turn to the case of company BBB, where the expected increase in volatility from the bet is also expected to raise the probability of costly lower-tail outcomes. In such a case, as we saw earlier, management should probably hedge to reduce the probability of financial trouble to acceptable levels. At the same time, however, top management should also consider subjecting its bets to an even higher standard of profitability to compensate shareholders for any associated increase in expected financial distress costs.

How much higher should it be? One method would be to assume that, instead of hedging, the firm raises additional equity capital to support the expected increase in volatility associated with the bet. In that case, the bet would be expected to produce the same risk-adjusted return on capital as the bet taken by company AAA, but on a larger amount of imputed "risk" capital.[21]

In sum, when devising a compensation scheme for those managers entrusted with making the firm's bets, it is critical to structure their incentive payments so that they are encouraged to take only those bets that are expected to increase shareholder wealth. Managers should not be compensated for earning average returns when taking larger-than-average risks. They should be compensated only for earning more than what their shareholders could earn on their own when bearing the same amount of risk.

This approach does not completely eliminate the problem discussed earlier caused by incentives for individuals to stand out in large organizations by taking risks. But traditional compensation schemes only reinforce this problem. If a risk-taker simply receives a bonus for making gains, he has incentives to take random bets because he gets a fraction of his gains while the firm bears the losses. Evaluating managers' performance against a risk-adjusted benchmark can help discourage risk-taking that is not justified by comparative advantage by making it more difficult for the risk-taker to make money by taking random bets.

CONCLUSION

This paper presents a theory of risk management that attempts to go beyond the "variance-minimization" model that dominates most academic discussions of corporate risk management. I argue that the primary goal of risk management is to eliminate the probability of costly lower-tail outcomes—those that would cause financial distress or

21. *The amount of implicit "risk capital" (as opposed to the actual cash capital)* backing an activity can be calculated as a function of the expected volatility (as measured by the standard deviation) of the activity's cash flow returns. For the distinction between risk capital and cash capital, and a method for calculating risk capital, see Robert Merton and André Perold, "Theory of Risk Capital for Financial Firms," *Journal of Applied Corporate Finance*, Vol. 6 No. 3 (Fall 1993). For one company's application of a similar method for calculating risk capital, see Edward Zaik et al., "RAROC at Bank of America: From Theory to Practice," *Journal of Applied Corporate Finance*, Vol. 9 No. 2 (Summer 1996). For a theoretical model of capital budgeting that takes into account firm-specific risks, see Kenneth Froot and Jeremy Stein, "Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach," Working Paper 96-030, Harvard Business School Division of Research.

make a company unable to carry out its investment strategy. (In this sense, risk management can be viewed as the purchase of well-out-of-the-money put options designed to limit downside risk.) Moreover, by eliminating downside risk and reducing the expected costs of financial trouble, risk management can also help move companies toward their optimal capital and ownership structure. For, besides increasing corporate debt capacity, the reduction of downside risk could also encourage larger equity stakes for managers by shielding their investments from "uncontrollables."

This paper also departs from standard finance theory in suggesting that some companies may have a comparative advantage in bearing certain financial market risks—an advantage that derives from information it acquires through its normal business activities. Although such specialized information may occasionally lead some companies to take speculative positions in commodities or currencies, it is more likely to encourage selective hedging, a practice in which the risk manager's view of future price movements influences the percentage of the exposure that is hedged. This kind of hedging, while certainly containing potential for abuse, may also represent a value-adding form of risk-taking for many companies.

But, to the extent that such view-taking becomes an accepted part of a company's risk management program, it is important to evaluate managers' bets on a risk-adjusted basis and relative to the market. If managers want to behave like money managers, they should be evaluated like money managers.