Risk Management, Capital Budgeting and Capital Structure Policy
for Insurers and Reinsurers[*]

Kenneth A. Froot[**]
Harvard University and NBER

November 2003

Abstract: This paper builds on Froot and Stein (1998) in developing a framework for analyzing the risk allocation, capital budgeting, and capital structure decisions facing insurers and reinsurers. The model incorporates three key features: i) value-maximizing insurers and reinsurers face product-market as well as capital market imperfections that give rise to well-founded concerns with risk management and capital allocation; ii) some, but not all, of the risks they face can be frictionlessly hedged in the capital market; iii) the distribution of their cashflows may be asymmetric, which alters the demand for underwriting and hedging. We show that these features result in a three-factor model that determines the pricing and allocation of risk and the optimal capital structure of the firm. This approach allows us to integrate these features into: i) the pricing of risky investment, underwriting, reinsurance, and hedging; and ii) the allocation of risk across all of these opportunities, and the optimal amount of surplus capital held by the firm.

I. Introduction

The cost of holding risk is a crucial concept for any corporation to understand. Most financial policy decisions, whether they concern capital structure, dividends, capital allocation, capital budgeting, or investment and hedging policies, revolve around the corporate costs of holding risk. These decisions can be made well only with a thorough understanding of how costly it is to originate and warehouse risk.

This issue is particularly acute for financial firms, since the origination and warehousing of risk constitutes their core value added. For these firms, capital is often their most expensive and important input in production. They deploy capital by holding a large number of financial risk positions that need to be evaluated. Moreover, these positions turn over and are competitively repriced far more often than the physical assets of non-financial firms. Unanticipated shocks to the demand for financial firms' products can be planned for and actively protected against, unlike those for non-financial firms. For these reasons, financial firms have the greatest need to understand the costs and benefits of holding risk.

Insurers and reinsurers certainly belong in this group. They have large underwriting and reinsurance portfolios, as well as investment and hedge portfolios. Often they have a large number of clients who view their insurance contracts as financially large and important claims.

However, there are two basic features of insurers and reinsurers that make them especially sensitive to the costs of holding risk. The first is that their customers – especially retail policyholders – face contractual performance risks which are large relative to their wealth. Once premiums are paid, policyholders worry whether their future policy claims will be honored swiftly and fully. Commercial bank customers, for example, are less exposed, because they are protected by deposit guarantees as depositors and are not exposed to bank performance risk as borrowers.

The large contractual performance risk borne by insurance and reinsurance customers makes demand sensitive to underwriter risk. It seems sensible that firms whose capital is at greater risk are likely, all else equal, to have a tougher time selling their products. There is considerable empirical evidence suggestive of this. For example, Sommer (1996) seeks evidence of an empirical relationship between insolvency risk and property-liability premiums. He finds that, all else

equal, firms with a greater ratio of surplus to assets have greater ratios of profit to premium.[1] In addition, he finds greater volatility of surplus-to-assets is associated with lower profitability rates. In addition, Grace, Klein, and Kleindorfer (2001, 2003) examine homeowners coverage in New York and Florida and find that higher-rated insurers get higher premiums. In particular, they find that rating is more important for premiums for customers who may have claims above state guarantee fund coverage limits than for those customers who would not have claims above state-specified coverage limits. Finally, Epermanis & Harrington (2001) find that insurers with high AM Best ratings grew annual premiums during 1992-1996 approximately 6% faster than those that were not highly rated. The effect is largest statistically and economically for changes in lower-rated and small firms, in the year of and year subsequent to the ratings change, and with respect to negative rating changes.[2]

There is thus considerable evidence that customer demand is sensitive to the financial standing of the insurer. Note that the change in customer demand may appear as either a change in the price an insurer can charge or a change in the quantity that the insurer can sell. The investigation of profits in Sommer (1996) and of premiums written in Epermanis and Harrington (2001) detects both of these channels.

While these studies suggest that customer demand falls with a decline in insurer rating, it is yet more difficult to marshal evidence that customer demand falls *by more* than capital provider demand with a decline in rating. Nevertheless, there are a number of reasons to believe that customers act as though they are more risk averse than capital providers, discounting future claims more heavily on the basis of even relatively small probabilities of failure. There are two kinds of explanations behind the greater product market sensitivity of customers, one is behavioral and the other is rational.

The behavioral story begins with Richard Zeckhauser's Russian roulette example. Suppose a randomly-spun revolver has one bullet in the chamber. How much would you pay to have it removed before pulling the trigger? How much, by comparison, would you pay to remove one

---

[1] This is not unrelated to the finding in nonfinancial corporations that past profitability is a positive predictor of equity capital as a percentage of total capital. See Myers (1984).

[2] Epermanis and Harrington (2001) are actually primarily concerned with the 'market discipline' hypothesis, which

bullet if the chamber for six bullets had four bullets in it? Zeckhauser both introspected and interviewed on this question. He emerged with the view that, contrary to what is predicted by Expected Utility Theory, most individuals say they would pay more to remove the last bullet than merely to reduce by one the number of bullets.[3] This example depicts a central idea in Kahneman and Tversky's Prospect Theory, one that they call the "certainty effect." Individuals value certainty, and will pay more to reduce uncertainty by one unit the lower is the probability of a bad outcome.

The behaviorist story is formalized in Wakker, Thaler, and Tversky (1997). They argue that "probabilistic insurance" – the name given by Kahneman and Tversky (1979) to an insurance contract that sometimes fails to pay the contractually legitimate claims of the insured – is deeply discounted by an expected utility maximizer who pays an actuarially-fair premium. They also provide survey evidence that the discount is striking in size: in comparison to a contract with no default risk, a contract with 1% independent default risk is priced 20%-30% lower by survey participants.

Of course, there is nothing about this story that is specific to insurance – the same will hold true of an individual's desire to eliminate the last bit of risk in a portfolio of securities. However, for multiple reasons, portfolios often are managed at relatively high levels of risk, suggesting that the price of risk reduction is low. Insurance claims, by contrast, have relatively low risk of performance failure, and therefore, the price of risk reduction should be higher. Sellers of consumer durables often stress this point, since a reputation for high quality is highly correlated with the belief that the product will not break.

The rational argument owes most to Merton (1993, 1995a, b). He argues that customers of financial firms value risk reduction more highly than do investors because customer costs of diversifying are much higher. In the case of insurance, considerable information transfer must take place between buyer and seller. Insurance contracts are complex instruments for buyers to understand and value. And, in the event of a claim, it is costly for the buyer to notify multiple insurers, coordinate their monitoring requirements, and negotiate settlements. All of these factors

---

states that firms that grow premiums quickly, but indiscriminately, receive subsequent rating downgrades.

[3] Richard Zeckhauser himself did not formally write up this example. However, see Kahneman and Tversky (1979) for the behavioralist view and Raiffa (1969) and Weinstein, Shepard, and Pliskin (1980) for the contrary expected-utility-theory view of the roulette game.

tend to make purchases of insurance lumpier and the exposure to a single firm's performance greater. Higher frictional costs of diversification therefore translate into effective risk aversion that is higher for customers than investors.

A related effect in insurance is that customer losses occur predominantly when customers' wealth is low and, therefore, the marginal value of wealth is high. If absolute risk aversion is declining in wealth (as many suspect it is), then a customer will be more averse to an insurer's failure to perform than to a debtor's failure to perform, even if the performance failure is of equivalent size.

A final aspect of the rational story is what might be called the 'hassle' factor – customers face hassles in dealing with a failed or failing insurer. Policies must be cancelled, claims and premium rebates must be pursued more vigilantly or even sacrificed, and new coverage must be found.[4] As the insurer's risk increases, the expected cost of these hassles is greater for customers than investors.

All of these are reasons to think that when insurer financial circumstances decline, customer demand not only falls, it falls *by more* than investor demand. There is some tentative evidence in the literature to support this view of 'excess' sensitivity. For example, Phillips, Cummins and Allen (1998) try to estimate directly price discounting for probability of insurer default. They find the discounting is about 10 times the economic value of the default probability for long-tailed lines, and 20 times for short-tailed lines. While this does not provide a direct comparison with price discounting in the capital markets, their numbers are large enough to suggest evidence of excess sensitivity. In any case, we emphasize this feature of insurance companies below in solving for optimal risk allocation, hurdle rates (pricing), and capital structure.[5]

The second feature that makes insurers and reinsurers especially sensitive to the costs of

---

[4] These deadweight costs may also need to be borne by investors in the aggregate. However, the marginal impacts for investors are lower both because institutional agents have dedicated capacity for these distress-related functions, and because the market-based trading of securities allows those with the lowest cost of performing such functions to acquire distressed instruments.

[5] There are a number of papers that rely on customer preferences to drive insurance pricing, some interacting with the capital markets and some not. See Zanjani (2002), Cummins and Danzon (1997), Cummins and Sommer (1996), Taylor (1994), and Hoerger, Sloan, and Hassan (1990).

holding risk is that they often face negatively asymmetric or skewed distributions of potential outcomes. Many insurance and reinsurance portfolios contain important exposures to catastrophic risks – natural and man-made events that create correlation in realized losses across large numbers of contracts and policies. Such events can be particularly damaging to insurers and reinsurers and can be expected to have first-order effects on risk allocation, pricing, and capital structure decisions.

Perhaps surprisingly, there has been relatively little work that distinguishes how changes in the shape of the distribution of outcomes affect these key decisions.[6] One reason for this is that there are an infinite number of higher-order moments, which makes working with them unwieldy. Another is that higher-order moments appear only rarely in the capital markets literature, in part because continuous-time finance allows for discrete time non-normality even though innovations are normally distributed. We adapt here a functional approach to describing asymmetric distributions, treating them as functional transformations of symmetrically-distributed normals. This allows us to generate general pricing and allocation results without having to specify an infinite series of moments. In this way, this paper takes a step toward explicitly incorporating asymmetrically distributed risks in formal corporate pricing and allocation metrics.

However, neither of the features of insurance and reinsurance firms we have highlighted – the sensitivity of customer demand and the asymmetric nature of underwriting exposures – themselves create a well-founded corporate concern with risk management and capital allocation. If companies can, at any time, costlessly access capital markets at fair value, then customer risk aversion and asymmetric distributions generally won't affect pricing and allocations. Nothing would stop the Modigliani Miller irrelevance theorems from applying.[7] For example, a firm could pursue a policy of massive overcapitalization – i.e., it could hold a very large amount of capital as an overly generous internal buffer – in order to drive its likelihood of default to zero. Clearly, we need something else, some form of capital-market imperfections that prevent such easy and costless

---

[6] Kraus and Litzenberger (1976) and Kozik and Larson (2001) add higher-order moments to the CAPM directly, the latter specifically in the case of insurance exposures. However, even if we were to use this for capital market's pricing, it does not adjust pricing for that which is relevant inside the firm. See the discussion in section III.D below.

[7] See Doherty and Tinic (1982) for an application of the Modigliani and Miller logic to insurance firms.

access to markets.[8]

For these imperfections, the approach we take here is to follow Froot and Stein (1998). As in that paper, we employ two kinds of capital market imperfections. The first is a 'carrying' cost of capital, i.e., equity capital held by an insurer or reinsurer is taxed while it is on the balance sheet. The implication is that, all else equal, an additional dollar of equity capital raises the market value of the firm by less than one dollar. The most straightforward rationale for this is corporate taxation, i.e., taxes that wouldn't be levied if the dollar were deposited instead in a mutual fund or other pass-through savings vehicle. Another rationale is agency concerns: large accumulations of internal capital raise the question of whether management will be able (or even intends) to deliver a high return. Either way, there is plenty of evidence to suggest that limits on the amount of external equity capital a single firm (especially a small one without any unique product-market breakthrough) can feasibly raise.

The second capital market imperfection we employ is an 'adjustment' cost of capital. This is the standard approach taken in models of financing under asymmetric information (see, for example, Myers and Majluf, 1984), where it becomes expensive for the firm to use external funds to add to existing internal capital. This adjustment-cost mechanism provides another motivation – in addition to the product-market and asymmetry concerns above – for the firm to avoid holding *too little* capital as a way of mitigating the carry costs. Following Froot and Stein (1998) and Froot, Scharfstein and Stein (1993), if the firm allows internal funds to run down too far, it will have to choose between cutting highly rewarding investments or incurring the high costs of external finance. Adjustment costs of capital would seem to be present in the insurance and reinsurance industries. For example, there is evidence that, in the aftermath of catastrophic events that deplete industry capital, adjustment costs can help to explain subsequently high prices and reduced availability of insurance and reinsurance.[9]

With these capital market imperfections in place, the firm has a well-founded concern with

---

[8] There are a large number of papers that help motivate the failure of Modigliani Miller, appealing to imperfections of various sorts. On the pricing of insurance and reinsurance claims, see, for example, Cummins and Phillips (2000), Garven and Lamm-Tennant (1997), Cagle and Harrington (1995), and Winter (1994). On capital (i.e., risk) allocation see Merton and Perold (1993), Froot and Stein (1998), and Myers and Read (2001).
[9] See, for example, Gron (1994a, b), and Froot (1997, 1999, 2001).

risk management and capital allocation from both product-market and capital-market tradeoffs. We combine these with the product-market sensitivity to risk and the asymmetry of exposures discussed above. As one might expect, both of these latter factors tend to make the insurer or reinsurer more conservative in taking on risk, more eager to diversify its investment and underwriting exposures, more aggressive in hedging, and more willing to incur carry costs of equity capital. As in Merton and Perold (1993) and Myers and Read (2002), our framework jointly and endogenously determines optimal hedging, capital budgeting and capital structure policies.[10]

We find that product-market considerations contribute additively to capital-market distortions in reducing the desire to hold risk and to price risk at fair market levels. We also find that product-market considerations strongly tilt the optimal level of capital and surplus toward higher levels. In addition, we find that the firm will explicitly price any distributional asymmetry in underlying payoffs. As might be expected, the costs of bearing negatively skewed exposures are generally much higher than those of positively skewed exposures. Furthermore, if the firm has negatively skewed exposures, it will seek more aggressive reinsurance and hedging and while its underwriting and investment strategies will be less aggressive and more diversified.

The remainder of the paper is organized as follows. Section II lays out the basic structure of the model. Second III analyzes the model's implications for risk allocation and pricing, i.e., the capital budgeting decision. Section IV then examines the capital structure issues. Section V provides more detailed interpretations of the results. Section VI shows how one can incorporate investment decisions in certain skewed distributions. Section VII concludes.

II. The Model: Timing and Assumptions

The model, which follows Froot and Stein (1998) closely in structure, has three time periods, 0, 1, and 2. In the first two periods, time 0 and 1, the insurer chooses its capital structure and then makes underwriting and hedging decisions. These two periods are at the heart of our analysis. They highlight the fact that consumers will pay less for an insurance contract written by a firm with low surplus with given risk. The last period is not absolutely essential. It closes the model by allowing

---

[10] See Venter and Major (2002) for a survey of techniques as applied to insurance.

the insurer to raise additional capital after paying (or defaulting on) its losses. At this stage additional frictional costs of raising capital can be added.

A. Time 0: insurer capital structure decision

At time 0, the insurer chooses how much equity capital, K, to hold. The capital goes toward financing various risks, in both investments and underwriting. There are, however, distortions that make the use of capital expensive, owing to we call deadweight costs to shareholders. These costs can arise from a number of sources. Perhaps the most concrete source would be income taxes. Corporate income taxes reduce the rates of return that K can provide to shareholders, because income is taxed at the corporate level before it comes to shareholders. A second source of deadweight cost is corporate opacity. It can create deadweight costs because shareholders cannot easily protect themselves if managers do not act to maximize shareholder value. Both taxes and opacity make it expensive for a firm to carry a large amount of capital. We summarize the deadweight costs of holding capital by τK, where τ is the effective "tax" rate. Naturally, in the absence of this tax, the accumulation of large equity capital balances would be costless, and the firm would have no motivation to limit its use of equity capital.

Also at time 0, we assume that the insurer inherits a given portfolio of risk exposures. This portfolio results in a time-2 random payoff $Z_P = \mu_P + \varepsilon_P$, where μ_P is a mean and ε_P is a mean-zero disturbance term. The risks arise from the insurer's pre-existing portfolio of investments in securities, derivatives, and underwriting risks. The total payoff from the insurer's internal funds in place as of time-0 is $Z_P + (1-\tau)K$. As discussed below, payoffs are realized in the model at time 2.

The payoff Z_P has components from both the capital markets and insurance (and reinsurance) market activities of the insurer. We define these components in such a way so that they are independent. To allow for explicit solutions above, we will also assume that the capital markets portion is normally distributed. We denote the capital market payoffs by $\mu_P^C + \varepsilon_P^C$, distributed normally with mean μ_P^C and variance $\left(\sigma_P^C\right)^2$.

The normality assumption is convenient for deriving explicit solutions below. But it is not very realistic, particularly for the insurance and reinsurance market positions, which may have strongly skewed payouts. Consequently, on the insurance market side, we allow the disturbance term to be non-normally distributed. The insurance portion of the payoff is given by $\mu_P^I + f(\varepsilon_P^I)$, where $f(\varepsilon_P^I)$, is a monotonic transformation of ε_P^I, a normally-distributed random variable with mean zero and variance $(\sigma_P^I)^2$. We assume that the function f is at least twice differentiable and that $f(\varepsilon_P^I)$ satisfies $E[f(\varepsilon_P^I)] = 0$ and $E[f'(\varepsilon_P^I)] = 1$, i.e., that the transformation of the underlying normal does not alter either the mean or the average scale of the distribution of ε_P^I.

Such a transformation is very flexible in describing option-like and other asymmetric payoffs based on a symmetrically distributed underlying variable. To describe what we might loosely call negatively-skewed payoffs, such as those on the underwriting book, a sufficient condition might be that $f' > 1$ for all $\varepsilon_P^I < 0$, and for positively-skewed payoffs such as those on a reinsurance portfolio, a sufficient condition would be that $f' < 1$ for all $\varepsilon_P^I < 0$.[11]

The mean payoff from the insurance market component, μ_P^I, also differs from the capital markets component. Specifically, we treat μ_P^C as given and unaffected by the insurer's decisions. For example, an insurer holding less capital, all else equal, will face the same capital-market investment opportunities as an insurer holding more capital. Think of the purchase of a futures contract, which provides the same expected return to any holder who has put up a pre-set amount of variation margin. Margin requirements have no investor-specific credit-sensitive component; indeed investors are anonymous to the exchange.

By contrast, we assume that μ_P^I contains an insurer-specific component. As the insurer gets riskier, its product-market investment opportunities degrade, i.e., its expected underwriting returns fall. This occurs even after the insurer has put up additional statutory surplus (margin) to

---

[11] There is no exact correspondence between the f function and the third moment of its distribution, skewness. The distribution of f defines all higher-order moments, not just the third. Thus, in the text below, we often refer to the distribution of f as negatively (or positively) *asymmetric*, rather than negatively (or positively) skewed, to be clear that we are not merely referring to the behavior of the third moment.

back any additional underwriting. Note that a decline in premiums that comes along with a decline in expected claims payments is *not* necessarily a decline in expected returns. When we say that product-market opportunities degrade, we mean that premiums must fall *by more* than expected insurance claims payments.

What matters for our model is that product market costs of capital rise faster than capital market costs of capital with additional risk, for given capital. The added degradation in mean underwriting return can occur for the reasons discussed above. First, there may be behavioral reasons – Kahneman and Tversky's Certainty Effect – that motivate individuals to pay relatively more to eliminate the last increments of uncertainty. Second, frictional costs of diversification suggest that customer exposures are proportionately larger than investor exposures, prompting greater effective risk aversion. Third, these frictional costs may include the 'hassles' customers must go through in the event that an insurer fails to perform.[12] And finally, there is the argument that the customer's marginal value of wealth conditional upon an insured loss is greater than the investor's marginal value of wealth conditional on the default of a similar sized investment.

For these reasons, we assume that expected product-market opportunities, μ_I, degrade with increases in the insurer's effective leverage. Capital-market opportunities, by contrast, do not degrade. As we will see, the company will take this into account in setting K initially. We represent product-market sensitivity by assuming that $\mu_P^I = \mu_P^I(s^2)$, where $s^2 = \left(\dfrac{\sigma}{K}\right)^2$, measures the squared standard deviation of company-wide risk per unit of capital. An increase in risk/capital – which is what we mean by 'effective leverage' – reduces the expected opportunity set, i.e., $\dfrac{d\mu_P^I}{d(s^2)} = \mu_P^{I'} < 0$.

To summarize the set-up at time 0: On the balance sheet, the insurer chooses an amount of

---

[12] We do not rule out the possibility that policyholders have their claims paid in some negative states of nature by third parties (guarantee funds, governments, etc.). Naturally, if these ex post payments occur without explicit risk-based ex ante charges, the insurer would have an incentive to seek risk through additional underwriting. In the paper, we abstract from the risk-seeking behavior that such ex post payment arrangements might incent. Regulatory authorities are charged with preventing this behavior, though probably not with complete success. See, for example, Hall (1999) for a discussion and evidence on the magnitude of the problem in the US.

cash capital, K, which is entirely equity-financed. Cash capital results in a deadweight cost of $K\tau$. This cost forms a wedge between the value of the assets inside the firm, and the value of the firm in the marketplace. The insurer also holds a portfolio of exposures with net payoffs of $Z_P = \mu_P + \varepsilon_P = \mu_P^C + \mu_P^I(s^2) + \varepsilon_P^C + f(\varepsilon_P^I)$, where the two components of ε_P, ε_P^C and ε_P^I, are defined so as to be independent, mean-zero normals. Because the premiums available in customer underwriting decrease with increases in firm-wide risk, $\mu_P^{I\,\prime} < 0$, the mean return on Z_P is endogenous to the choice of K.

Notice that the deadweight costs of capital thus far come from the stock of capital, K, and not from any costs of adjusting the capital structure. This is different from most standard models of financing under asymmetric information. In those models, costs are generally incurred when raising incremental external funds, not by having equity capital on the balance sheet per se. Indeed, we assume below that there are exactly such flow costs of new external finance at time 2.

The absence of any cost to raising capital at time zero is clearly a shortcut. Nevertheless, it allows us to focus better on the appropriate long-run "target" level of capital for the insurer. Over long periods of time (over which adjustment costs can best be amortized), what kind of ratings standard is best for the insurer or reinsurer? How should it position itself in that regard? We recognize that, if at any point in time, the insurer is far away from its ideal target, it may face costs of adjustment in getting to the target quickly, but it is nonetheless interesting to ask the question of what the target should be.

B. Time 1: insurer underwriting and hedging decisions

At time 1, the insurer makes additional decisions. It decides how much new investment and underwriting to undertake and how much hedging or reinsurance to invoke. We treat investment, underwriting, hedging, and reinsurance in a single new-opportunity specification. As above, the jth new opportunity has an expected and unexpected component, given by $Z_{N,j} = \mu_{N,j} + \varepsilon_{N,j}$, where $\mu_{N,j}$ is the mean payoff, and $\varepsilon_{N,j}$ is a mean-zero normally-disturbance term. Each opportunity can represent a new underwriting opportunity, a reinsurance contract, a capital-markets hedging vehicle,

or any other self-financing new investment.

   As above, it is useful to think of each opportunity as being comprised of two independently-distributed parts, a capital-market and an insurance-market component. In what follows, we treat the capital-market component as fully liquid and costlessly traded. As a result, it will always trade at 'fair' prices. We also assume that the insurance market exposures are illiquid, and allow for the possibility that they are mispriced, offering returns that are not fair based on their capital market risks. This latter group of exposures could easily include not only traditional insurance and reinsurance risks, but also credit exposures. Indeed, any risk with a component orthogonal to the major capital markets risks could be considered as an insurance market exposure rather than a capital market exposure. All that matters for our analysis is that the market-wide factors priced by the capital markets trade costlessly and at fair prices.

   To see all this more precisely, we additively decompose the total payoff from the jth new opportunity into the independent capital and insurance market exposures:

$$Z_{N,j} = \mu_{N,j} + \varepsilon_{N,j} = \left(\mu_{N,j}^{I} + \mu_{N,j}^{C} \right) + \left(\varepsilon_{N,j}^{I} + \varepsilon_{N,j}^{C} \right),$$

where the superscripts I and C denote, respectively, insurance and capital market exposures. We assume that the jth new-opportunity disturbances, $\varepsilon_{N,j}^{I}$ and $\varepsilon_{N,j}^{C}$, are independently distributed mean-zero normals. (Normality is needed here in order to derive explicit solutions for the choice variables. However, our approach allows for more generality in the distributions of the disturbances and we comment on this below.) The combined disturbance, $\varepsilon_{N,j}$, may contain any combination of insurance and capital market components.

   The magnitude of the insurer's exposure to the new opportunity is a choice variable, given by n_j. Examples of n_j would be the number of policies written, or the number of futures or reinsurance contracts used to hedge, or the number of credit default swaps purchased or written, etc., even a new business opportunity can be described in this way. In other words, n_j is meant to cover essentially any new risk decision that the firm contemplates. Thus, the total payoff from the new underwriting, hedging, and investment opportunities is $nZ_N = n\left(\mu_N + \varepsilon_N \right)$

13

$$= \sum_j n_j \left(\mu_{N,j}^I + \mu_{N,j}^C + \varepsilon_{N,j}^I + \varepsilon_{N,j}^C \right).$$

The insurer's realized internal wealth at time 2, w, is therefore given by:

$$w = Z_P + nZ_N + K(1-\tau) \tag{1}$$

In words, the amount of cash the insurer has on hand at time 2 to pay claims will depend on the realizations on its pre-existing capital market and insurance exposures, Z_P, on the extent of its new underwriting, investment, and hedging outcomes, nZ_N, and on the amount of capital, K, raised at time 0.[13]

The sensitivity of underwriting opportunities to firm-wide risk levels creates a meaningful need for risk management for the value-maximizing insurer. To see this, suppose the insurer finds an underwriting opportunity where, loosely speaking, premiums are high relative to risk. In this circumstance, the insurer will choose to underwrite, setting $n_j > \tilde{0}$. However, as n_j grows, internal risk increases, raising s^2 and reducing premiums on *all* existing underwriting by $\mu_P^{I\,\prime} d(s^2)$. All else equal, this product-market externality lowers expected internal funds, $E[w]$, and raises internal hurdle rates. The underwriting opportunity itself and other risky investments appear less attractive, while costly hedging opportunities and balance-sheet capital appear more attractive. The sensitivity of product market opportunities to firm-wide risk therefore creates a well-founded motivation for insurer risk management.

Lastly, we note that this specification of $\mu_P^I = \mu_P^I(s^2)$ need not be thought of as restricted to existing underwriting business. Indeed, it is likely that higher s^2 would degrade future product market opportunities, and not simply underwriting already in place. In a sense, s^2 and its impact on expected underwriting performance can be viewed as partly coming through a reputational

---

[13] Note that in our specification, the insurer's risk level impacts expected internal funds only through $\mu_P(s^2)$; μ_N is not a similar function of squared risk per unit of capital. One might expect, of course, that additional risk would affect new opportunities as well as the existing opportunity set. Including this in the specification would not change any of the results importantly. We leave out the sensitivity of the new opportunity to risk for simplicity; it does not change the basic results.

channel. Thus, if these effects – both on business in place and on future business through reputation – are important, then the magnitude of $\mu_P^I d(s^2)$ may be large if measured as a fraction of business that is already in place.

C. Time 2: the realization of cashflow, and the insurer response

Next we need to specify how the insurer uses internal funds to create value. The simplest possibility would be to assume that the insurer terminates its operations at time 2, and immediately dividends back all realized internal funds, w, to investors. If the insurer were to do this then the *ex-post* value of the firm, which we define by P, would simply be equal to the realization of w, so that $P=w$. Even with such a simple dividend assumption, insurers would have an incentive to control the risk of their internal funds. A value-maximizing insurer would seek, all else equal, to control non-market risk in order to increase expected underwriting returns, thereby increasing $E[P(w)]$.

However, by controlling the risk of their internal funds, insurers and reinsurers are also likely to face better fund raising opportunities in the *capital* markets. That is, with lower volatility of internal capital, the cost of raising external funds is likely to be lower. This is the theme in much of the literature on costly external capital markets (see, for example, Greenwald, Levinson, and Stiglitz, 1991). Here we follow Froot, Scharfstein and Stein (1993), by assuming that after w is realized, the company has a further business opportunity – e.g., it might be able to open a new line of business, buy an existing business, etc., in addition to needing its capital to fund its ongoing book of business. This investment requires a cash commitment of I, and yields a gross return of $H(I)$, where $H(I)$ is an increasing, concave function of I. The investment can either be funded out of internal sources, or funds can be raised externally in an amount e. Thus $I = w + e$. The critical point is that there are convex costs to raising external finance, given by $C(e)$. This means that it becomes more costly to raise funds the larger is the amount that must be financed externally.[14,15]

---

[14] Froot, Scharfstein and Stein (1993), give a number of microeconomic rationales – based on agency and/or information problems – to justify this sort of specification for the $C(E)$ function. They show how this convex functional form arises in a specific optimal contracting setting, a variant of the costly-state verification model due to Townsend (1979) and Gale-Hellwig (1985). Stein (1996) generates a similar formulation in a banking model where non-deposit liabilities are

Denote by $P(w)$ the solution to the insurer's time-2 problem:

$$P(w) = \max_I H(I) - I - C(e), \text{ subject to } I = w + e. \tag{2}$$

FSS demonstrate that $P(w)$ can be rigorously derived in the context of one standard optimal contracting models as an increasing concave function. Thus, $P_w > 1$ and $P_{ww} < 0$.

The concavity of the $P(w)$ function with respect to w generates a second rationale for insurer risk management. This concavity in turn arises from two sources. First, is the convexity of $C(e)$, which matters to the extent that fluctuations in internal cash result in fluctuations in additional costs of raising external funds. Second is the concavity of $H(I)$, which matters to the extent that fluctuations in internal cash result in fluctuations in investment, lowering the average return on investment. Loosely speaking, the more difficult it is for the insurer to raise external funds on short notice at time 2, the more averse it will be to fluctuations in its time-2 internal wealth w. Thus, fund raising opportunities in the capital markets become less appealing as internal funds become increasingly unpredictable.

The derivative P_w summarizes the ex post value of having an additional dollar of internal funds. The value $E[P_w]$ summarizes the ex ante value of this dollar. Naturally, a property of the model solution must be that $E[P_w] \geq 1$: an additional dollar of internal wealth has expected ex ante value of at least one dollar. To see this, note that even if there were no costs of carrying capital ($\tau = 0$), a value-maximizing firm would never raise a dollar of external funds to increase firm value by less than a dollar. In the presence of costs of carrying capital ($\tau > 0$), the inequality will be strict, i.e., $E[P_w] > 1$.

---

subject to adverse selection problems.

[15] It is straightforward to allow the insurer's gross return on investment to be sensitive to product-market perceptions about time 2 risk and time 0 capital. For example, we could specify that $H = H(I, s^2)$, where $H_s < 0$, $H_{ss} > 0$. This would make more explicit the reputational channel through which past risk and capital behavior affect future product market opportunities.

16

III. Analysis

To understand the properties of the model, it is best to work backwards. We have already seen that any given realization of w at time 2 can be mapped into a non-stochastic payoff, $P(w)$. Next we ask from the perspective of time 1, when w is uncertain, what amount of new investment, underwriting, hedging and reinsuring will maximize market value? From there we proceed backward again, solving for the right amount of capital given the costs of carrying it at time 0, the costs of having too little capital at time 1, and the costs of relying on external financing to replenish capital at time 2.

A. Shareholder valuation of the insurer at time 1

From the perspective of time 1, the shareholder payoff, $P(w)$, is a random variable. In order to value the payoff, we need a pricing model. We assume that the company's stock is like the market-wide capital market exposures, in the sense that the shares trade costlessly and are priced fairly. Specifically, we assume that the fair pricing rule has required returns as a linear, decreasing function of covariance with a 'market' factor, M. It is straightforward to generalize this to a multifactor setting, where covariances with many factors determine fair-market required returns. With just a single factor, the value of the insurer's shares V, will be:

$$V = \left(E[P(w)] - \gamma \operatorname{cov}(P(w), M)\right)/(1+r) \tag{3}$$

where r is the riskless rate of interest between time 1 and time 2.

The first thing we can do with this pricing model is value the pure capital market exposures that are contained in both Z_P and Z_N. Expected returns for these exposures, μ_P^C and μ_N^C, will be determined by the pricing model in equation (3), because liquid exposures are priced fairly and have no transaction costs. Fair pricing implies simply that $\mu_P^C = \gamma \operatorname{cov}(\varepsilon_P^C, M)$ and $\mu_N^C = \gamma \operatorname{cov}(\varepsilon_N^C, M)$.

Pricing these liquid exposures also helps clarify our assumption about the sensitivity of mean

underwriting opportunities to risk per unit of capital, $\mu_P^I = \mu_P^I(s^2)$. Fair market pricing would say that μ_P^I should be a function solely of covariance with the market factor, not a function of total firm-wide risk. If policyholders were to behave exactly like investors, we would expect their required returns to increase with the market risk, not with the total risk, of the insurer's portfolio. In other words, as total risk increases, the insurance market opportunity set degrades faster than the capital market opportunity set because the latter responds only to the *systematic* risk of the firm.

Substituting the definition of fair pricing into equation (1), we can rewrite time 1 internal funds as:

$$w = \gamma \operatorname{cov}\left(\varepsilon_P^C + \sum_j n_j \varepsilon_{N,j}^C, M\right) + \mu_P^I(s^2) + \sum_j n_j \mu_{N,j}^I + \varepsilon_P^C + f(\varepsilon_P^I) + \sum_j n_j \varepsilon_{N,j} + K(1-\tau),$$

(4)

where μ_P^I and $\mu_{N,j}^I$ can be interpreted as expected excess returns, above the fair market, earned on insurance market exposures. Note that we have used the fact that the insurance market shocks are independent of M, thereby setting their covariances to zero. To collapse terms, it is useful to rewrite (4) as:

$$w = \mu + \varepsilon + f(\varepsilon_P^I),$$

(5)

where $\mu \equiv \gamma \operatorname{cov}\left(\varepsilon_P^C + \sum_j n_j \varepsilon_{N,j}^C, M\right) + \mu_P^I(s^2) + \sum_j n_j \mu_{N,j}^I + K(1-\tau)$ and $\varepsilon \equiv \varepsilon_P^C + \sum_j n_j(\varepsilon_{N,j})$.

Equation (5) has the virtue of separating w into a mean plus a normal disturbance and non-normal disturbance.

B. Optimal hedging policy at time 1 for capital market exposures

The insurer designs its risk management policy so as to maximize shareholder value, V. Using equation (3), this means it will choose n_j at time 1 to satisfy:

$$\frac{dV}{dn_j} = \frac{dE[P(w)]}{dn_j} - \gamma \frac{d \operatorname{cov}(P(w), M)}{dn_j} = 0, \tag{6}$$

where all expectations are taken with respect to the ε's, and where dn_j represents a change in the quantity of the jth new investment. Taking derivatives and using the definition of covariance, we can write:

$$\frac{dV}{dn_j} = E[P_w]E\left(\frac{dw}{dn_j}\right) + \operatorname{cov}\left(P_w, \frac{dw}{dn_j}\right) - \gamma \frac{d \operatorname{cov}(P(w), M)}{dn_j}. \tag{7}$$

To simplify this, we apply the definition of internal funds from equation (4), adding the fact that $\dfrac{d(s^2)}{dn_j} = \dfrac{d \operatorname{var}(\varepsilon + f(\varepsilon_P^I))}{K^2 dn_j} = \dfrac{2}{K^2} \operatorname{cov}(w, \varepsilon_{N,j})$ to derive:

$$\frac{dw}{dn_j} = \mu_P^I \cdot \frac{2}{K^2} \operatorname{cov}(w, \varepsilon_{N,j}) + \mu_{N,j}^I + \gamma \operatorname{cov}(\varepsilon_{N,j}^C, M) + \varepsilon_{N,j}. \tag{8}$$

We can then rewrite equation (7) as:

$$\frac{dV}{dn_j} = E[P_w]E\left(\frac{dw}{dn_j}\right) + \operatorname{cov}\left(P_w, \varepsilon_{N,j}\right) - \gamma \frac{d \operatorname{cov}(P(w), M)}{dn_j}. \tag{9}$$

To simplify this, we need to express the covariances in terms of the underlying random variables. Note first that, by assumption, both $\varepsilon_{N,j}$ and M are normally distributed. This allows us to use a generalization of Stein's Lemma (1981) for normal random variables x, y, and z:

$$\text{cov}(g(x,y),z) = E[g_x]\text{cov}(x,z) + E[g_y]\text{cov}(y,z), \tag{10}$$

where g is a continuous function (subject to some mild regularity conditions) and the expectations are taken with respect to the joint distribution of x and y.

Using this fact, equation (8), and simplifying, we can rewrite equation (9) as

$$\frac{dV}{dn_j} = E[P_w]\left(\mu_P^{I\prime}\frac{2}{K^2}\text{cov}(w,\varepsilon_{N,j}) + \mu_{N,j}^I\right) + E[P_{ww}]\text{cov}(\varepsilon,\varepsilon_{N,j}) + E[P_{ww}f']\text{cov}(\varepsilon_P^I,\varepsilon_{N,j})$$
$$- \gamma E\left[P_{ww}\frac{dw}{dn_j}\right]\text{cov}(w, M).$$

$$\tag{11}$$

Next, assume that there are in total $J+2$ new opportunities. The first J describe opportunities which contain combinations of insurance and capital market exposures. The last 2 are reserved for capital market exposures that span all the capital market disturbances.[16] With this in mind, suppose that the $J+2^{nd}$ new product market opportunity is a costlessly-traded instrument with payoffs identical to the market return, M. Since capital market opportunities are fairly priced and there is no insurance market component, we have that $\varepsilon_{N,J+2} = \varepsilon_{N,J+2}^C = M$ and $\mu_{N,J+2}^I = 0$. The independence of insurance and capital market components means that $\text{cov}(\varepsilon_P^I,\varepsilon_{N,J+2}) = 0$. Thus, with the jth new opportunity equated a linear investment in the market-wide risk, M, equation (11) simplifies to:

---

[16] We can add arbitrarily many of such factors, but with no change in generality or important results.

$$\frac{dV}{dn_j} = E[P_w]\left(\mu_P^{I\prime}\frac{2}{K^2}\mathrm{cov}(w,M)\right) + E[P_{ww}]\mathrm{cov}(w,M) - \gamma E\left[P_{ww}\frac{dw}{dn_j}\right]\mathrm{cov}(w,\,M) = 0, \qquad (12)$$

where we substituted M for $\varepsilon_{N,J+2}$, and $\mathrm{cov}(w,M)$ for $\mathrm{cov}(\varepsilon,\varepsilon_{N,J+2})$.

We can then state our first proposition, taking as given the first $J+1$ new opportunity decisions, n_j:

Proposition 1: It is optimal for the firm to choose an amount of market-risk hedge such that $\mathrm{cov}(w,M)=0$. This implies that the optimal hedging amount is the minimum variance

hedge ratio, i.e., $n_{J+2}^* = \dfrac{-\mathrm{cov}\left(\varepsilon_P^C + \displaystyle\sum_{j=1}^{J+1} n_j \varepsilon_{N,j}^C , M\right)}{\mathrm{var}(M)}$.

To prove this, one can see readily that equation (12) is satisfied by setting n_J such that $\mathrm{cov}(w,M)=0$.[17] In other words, the insurer will hedge out all of the market risk from internal funds, w, minimizing its variance with respect to M. The firm's market value rises because the reduction in the variance of internal funds improves product market opportunities and reduces the expected costs of future external finance. The optimal hedge ratio is given by solving $\mathrm{cov}(w,M)=0$ for n_{J+2}, with n_{J+2}^* above as the solution. In words, it will always be optimal to strip market risk entirely from the fluctuations in internal funds. Any exposure to M contained in the pre-existing portfolio, or in other new investment opportunities, is offset with the $J+2^{nd}$ hedge instrument, with returns equal to M.

Our next result is analogous: like the market risk, M, other capital markets exposures will be fully hedged, provided that they are fairly priced. To see this, let the $J+1^{st}$ new opportunity be a hedge of all other capital market exposures that are independent of M. There is no insurance component to the $J+1^{st}$ new opportunity, so that $\varepsilon_{N,J+1} = \varepsilon_{N,J+1}^C$, $\mu_{N,J+1}^I = 0$, and

---

[17] We assume that the second-order conditions are satisfied throughout.

$\text{cov}\left(\varepsilon_P^I, \varepsilon_{N,J+1}\right) = 0$. This and the previous hedge of M implies that the last two terms of equation (11) disappear.

Taking the first $J+1$ new investment decisions as given, it is then easy to show:

Proposition 2: the optimal hedge ratio for other capital market exposures sets internal funds to be uncorrelated with the hedge of capital market exposures, $\varepsilon_{N,J+1}^C$. This implies that the hedge is again set such that $\text{cov}\left(w, \varepsilon_{N,J+1}\right) = \text{cov}\left(\varepsilon, \varepsilon_{N,J+1}\right) = 0$, with the optimal amount corresponding to the minimum-variance hedge amount, $n_{J+1}^* = \dfrac{-\text{cov}\left(\varepsilon_P^C + \sum_{j=1}^{J} n_j \varepsilon_{N,j}, \varepsilon_{N,J+1}\right)}{\text{var}(\varepsilon_{N,J+1})}$.

Note that since these other capital markets exposures are assumed to be orthogonal to M, the solution for the $J+1^{\text{st}}$ hedge does not depend on n_{J+2}^*.[18]

Overall, the hedging of capital market risks strips out those exposures entirely from internal funds. Thus, the expression for internal wealth with capital market hedges in place becomes:

$$w_H = \mu_P^I(s^2) + \sum_j n_j \mu_{N,j}^I + f(\varepsilon_P^I) + \sum_j n_j \varepsilon_{N,j}^I + K(1-\tau), \tag{13}$$

where the optimal capital market hedges have negated all residual capital markets exposures from equation (4), $\gamma \text{cov}\left(\varepsilon_P^C + \sum_j n_j \varepsilon_{N,j}^C, M\right) + \varepsilon_P^C + \sum_j \varepsilon_{N,j}^C$.

These results echo those in Froot and Stein (1998): fairly priced exposures will be hedged fully by the firm. By raising the variability of internal funds, the costs that such risks

---

[18] In general, if the capital market exposures are not independent, then the relevant n_j first-order conditions must be

impose on the firm will not be fully compensated by the fair market expected return.

C. Optimal hedging policy at time 1 for illiquid exposures

With the capital market risks fully hedged at fair market prices, we are ready to add in new opportunities that contain insurance market exposures, i.e., those portions of risk that may be illiquid. There are two things that distinguish the required return and hedging decisions of these illiquid components from the capital market counterparts that we examined above. The first is that the insurance exposures interact with the non-normal payoffs already in the portfolio. This means that insurance instruments may be able to alleviate the costs, or enhance the benefits, of the asymmetrically distributed risks already in the firms' portfolio. The second is that these exposures are illiquid, and therefore may have required returns that differ from fair market returns.

We explore the implications of these first for the optimal amount of investment in the new opportunity. We then look to the hurdle rates that are implied, to understand better the firm's reservation price for new opportunities. We do this below for the jth new opportunity, which, unlike in the previous subsection, may contain both capital market and insurance market exposures.

To see the implications of these, we rearrange equation (11), where the last term has disappeared due the full hedging of capital market exposures.[19] We then have one of our two main propositions in the paper:

Proposition 3: The optimal amount of the new opportunity is given by three factors: the excess risk-adjusted return relative to its own variance, adjusted by the firm's tacit risk aversion,

$\left(\dfrac{1}{F+G} \right)\left(\dfrac{\mu_{N,j} - \gamma\,\mathrm{cov}(\varepsilon_{N,j}^{C}, M)}{\mathrm{var}(\varepsilon_{N,j}^{I})} \right)$, a kind of asymmetry-adjusted covariance of the opportunity

---

solved simultaneously rather than sequentially as above. See the discussion below.

[19] We also use the fact that with $E[f'] = 1$, $\mathrm{cov}\!\left(f(\varepsilon_{P}^{I}), \varepsilon_{N,j}^{I}\right) = \mathrm{cov}\!\left(\varepsilon_{P}^{I}, \varepsilon_{N,j}^{I}\right)$.

relative to own variance, $\left(\dfrac{\tilde{G}}{F+G}\right)\left(\dfrac{\text{cov}(\varepsilon_P^I,\varepsilon_{N,j}^I)}{\text{var}(\varepsilon_{N,j}^I)}\right)$, and the minimum-variance amount of the

opportunity, $\dfrac{\text{cov}(w_j,\varepsilon_{N,j}^I)}{\text{var}(\varepsilon_{N,j}^I)}$:

$$n_j^* = \left(\frac{1}{F+G}\right)\left(\frac{\mu_{N,j} - \gamma\,\text{cov}(\varepsilon_{N,j}^C, M)}{\text{var}(\varepsilon_{N,j}^I)}\right) - \left(\frac{\tilde{G}}{F+G}\right)\left(\frac{\text{cov}(\varepsilon_P^I,\varepsilon_{N,j}^I)}{\text{var}(\varepsilon_{N,j}^I)}\right) - \left(\frac{\text{cov}(w_j,\varepsilon_{N,j}^I)}{\text{var}(\varepsilon_{N,j}^I)}\right),$$

where $w_j = w - n_j\varepsilon_{N,j}^I$ is total internal funds less the disturbance from the jth new opportunity.

Before we interpret the proposition in its entirety, several terms in the above expression require definition. First, $F = -\mu_P^{I\,\prime}\dfrac{2}{K^2} > 0$ describes the sensitivity of internal funds to an increase in firm-wide risk that comes through the product-market channel. Essentially, F measures the magnitude of the negative demand externality created by customers who are more sensitive to diversifiable risk than are capital providers. This externality is greatest when leverage is high; that is, when the variance of internal funds is large relative to external capital. If the capital base grows, but firm-wide risk remains constant, $\mu_P^{I\,\prime}$ (and therefore F) moves toward zero. The hypothesis is that customers are concerned about risk not absolutely, but relative to the size of the capital base.[20]

The second term, $G = \dfrac{-E[P_{ww}]}{E[P_w]} > 0$ measures the concavity of firm-wide valuation in the capital markets (see Froot and Stein, 1998). It is similar to absolute risk aversion based on an investor's utility function and is the result of variability in expected future costs of raising external finance. The variability of internal funds not only negatively impacts product-market

---

[20] We abstract from any absolute size effects, which would have customers be more sensitive to a given risk/capital ratio for smaller firms. There is probably a strong argument that such size effects can be important, in that there are at least some fixed costs of producing insurance.

opportunities through F; it also degrades future growth opportunities by making external funds on average more expensive. G is a declining function of the realization of w (F is actually a function only of the ex ante properties of w). G therefore converges toward zero as w becomes large relative to K.

The terms F and G appear in an additive way because they are complementary channels through which changes in firm-wide risk impacts value. To form a measure of the effective risk aversion of the insurer, both F and G are required. If $F = G = 0$, then the insurer acts as though it is risk neutral, being willing to undertake an infinite amount of a new investment with positive risk-adjusted excess return, $\mu_{N,j} - \gamma \operatorname{cov}(\varepsilon_{N,j}^C, M)$. When $F + G > 0$, the insurer's response to a positive NPV new opportunity is more muted, as long as there is some unhedgeable risk imposed on firm-wide capital as a result (i.e., $\operatorname{var}(\varepsilon_{N,j}^I) > 0$).

The analogy with risk aversion is not exactly correct for two reasons. First, F and G are endogenous to the properties and realization of w; in many cases investor risk aversion is treated as an exogenous parameter, determined entirely by the form of the utility function. Second, and more importantly, F and G measure the willingness of the insurer to pursue risky returns in excess of *fair value*. By contrast, investor risk aversion measures the willingness to pursue risky returns in excess of the risk free rate. An investor will always want to hold at least a small amount of any risk that has positive expected returns above the risk free rate. A corporation will not wish to hold any risk unless it at least pays a return greater than fair value as determined in the capital markets.

Third, $\widetilde{G} = \dfrac{(E[P_{ww}] - E[P_{ww}f'])}{E[P_w]}$ measures the impact of the payoff asymmetry of pre-existing exposures. To see this, suppose that $f(\varepsilon_P^I)$ describes the negatively asymmetric payoffs of a typical insurance contract – occasional large negative outcomes offset by far more frequent small positive outcomes. In that case, $E[f'] > 1$ for $\varepsilon_P^I < 0$ and $E[f'] < 1$ for $\varepsilon_P^I > 0$. Given the concavity of $P(w)$, this creates the presumption that $\widetilde{G} > 0$. Loosely speaking, the more negatively asymmetric is the pre-existing portfolio, the more concave is the value function in expectation, i.e., as the distribution of $f(.)$ becomes more negatively asymmetric, $E[P(w)]$

falls. A positively asymmetric risk will provide the firm the additional benefit of reducing the concavity of the value function, so that $\widetilde{G} < 0$. In this way, the asymmetry of pre-existing payoffs enhances or reduces the potency of individual new opportunities, depending how they interact with the underlying risk, ε_P^I. A firm will want to hedge to a greater (lesser) extent to reduce the negative (positive) asymmetry of its underlying payoffs.

Indeed, Proposition 3 shows that the firm will hedge a negatively asymmetric exposure *beyond* the minimum variance amount. To see this, consider a new hedging opportunity that pays no return above fair market value, $\mu_{N,j} - \gamma \operatorname{cov}(\varepsilon_{N,j}^C, M) = 0$, so that the first term in Proposition 3 disappears. If the pre-existing positions all have normal distributions, the firm will of course want to hedge by an amount that eliminates the pre-existing exposure to the risk, $\dfrac{\operatorname{cov}(w, \varepsilon_{N,j}^I)}{\operatorname{var}(\varepsilon_{N,j}^I)}$. However, if its pre-existing payoffs are non-normal, it will wish to go further, hedging more (less), provided that by doing so it can reduce the negative (increase the positive) asymmetry in its pre-existing exposures. The ability to use the hedge to reduce the negative asymmetry in payoffs is given by $\widetilde{G} \operatorname{cov}(\varepsilon_P^I, \varepsilon_{N,j}^I)$.

We can take this one step further by looking back to the first-order condition in Equation (9) above. It shows that the firm is actually trying to reduce covariance of each exposure with the shadow value of internal funds, P_w. A perfect set of hedges would completely eliminate uncompensated covariation in P_w. For normally distributed risks, this is equivalent to eliminating the exposure with internal funds, w, itself. However, if the distribution of pre-existing exposures is non-normal, the asymmetry generally interacts with the concavity of $P(w)$. In that case, minimizing the variance of P_w is no longer the same as minimizing the variance of w. The optimal n_j must take into account the new investment's impact on the covariation of P_w by altering the asymmetry of w.

Finally, note that, like F and G, $\widetilde{G}$ falls to zero as internal funds increase relative to initial capital. Asymmetry in the pre-existing payoff distribution is less costly when there is plenty of internal financial slack. Another way to say this is that, when w is large, the shadow

value of internal funds, P_w, converges toward 1, and there is little need to manage internal funds. As a result, F, G, and $\widetilde{G}$ are all near zero.

Now that we have dissected the optimal new opportunity decision rule, it is straightforward to rearrange the expression to derive new opportunity hurdle rates:

Proposition 4: The required incremental rate of return on an incremental amount of the jth new opportunity is given by:

$$\mu_{N,j} = \gamma \operatorname{cov}(M, \varepsilon_{N,j}^C) + (F+G)\operatorname{cov}(w, \varepsilon_{N,j}^I) + \widetilde{G}\operatorname{cov}(\varepsilon_P^I, \varepsilon_{N,j}^I),$$

Proposition 4 shows that pricing internal risks requires a three-factor model. The first factor is standard – the quantity of market risk, $\operatorname{cov}(M, \varepsilon_{N,j}^C)$, multiplied by the price of market risk, γ.

The second factor is the quantity of firm-wide risk, $\operatorname{cov}(w, \varepsilon_{N,j}^I)$, multiplied by the price of firm-wide risk. Note that firm-wide risk is the second factor in the Froot and Stein (1998) two-factor model. This source of risk is also key in related studies, such as that of Myers and Read (2001) and Merton and Perold (1993). However, in the present context, the price of firm-wide risk, $F+G$, is greater than in previous studies. That is because of the additional product-market channel through which firm-wide risk impacts firm value. Notice also that the quantity of firm-wide risk is measured with respect to only the insurance market component of the jth new opportunity. Because the capital market risks are fully and costlessly hedged, there is no need to penalize a new opportunity for containing them.

The third factor is due to the asymmetry of firm-wide payoffs. It is the product of the quantity of asymmetry risk, $\operatorname{cov}(\varepsilon_P^I, \varepsilon_{N,j}^I)$ – i.e., the covariance of the jth opportunity with the asymmetrically distributed components of internal funds – with the price of asymmetry risk, $\widetilde{G}$. It is worth reemphasizing that our definition of asymmetry does not correspond precisely with skewness, or any particular group of moments of the asymmetric payoff distribution. The price

of asymmetry risk comes not from asymmetry per se, but from the interaction of the asymmetric distribution with variation in the marginal value of internal funds, P_w.

Of course, for both empirical and theoretical reasons, it is sensible to assume that an insurer's or reinsurer's pre-existing portfolio will be negatively asymmetric. This creates the presumption that $\widetilde{G}$ will be positive for these firms, and probably for financial intermediaries in general.

D. <u>Commentary on the pricing of asymmetric distributions</u>

It is often argued that much of the standard valuation models in finance are poorly suited for incorporating the pricing of asymmetric payoffs. For example, because investors (and firms) dislike negative skewness, capital market pricing formulae might include a factor based on the skewness or co-skewness of a position. Others in finance rebut this criticism, however, by allowing for continuous-time liquidity and trading with changing first and second moments. In such a case, negatively skewed distributions over discrete time intervals can often be replicated by Brownian motions (distributed normally), with conditional expected returns and variances that evolve over time. Basically, the effects of asymmetries on pricing can often be reproduced by more frequent trading.[21]

In our setup, we give the benefit of doubt to the continuous trading argument for a company's equity. (Although for insurance companies, in particular, the possibility of a large, sudden event that discretely diminishes value, like that of an earthquake or terrorist attack, is very real.) Our valuation function, equation (3), therefore ignores distributional asymmetries in determining the value of an asset based on expected cash flow and its covariance with the market.

However, even after ignoring return asymmetries for the pricing of cashflows in equation (3), asymmetries matter for determining the deadweight costs facing the firm. Low realized levels of

---

[21] This assumes that over the discrete interval, the value path is continuous, so that every value path can be drawn without "lifting the pencil." Gaps in the value path may indeed require additional factors to be priced accurately. Notice, however, that while individual insurance contracts or derivatives may indeed have such value gaps, if each of these is small and reasonably independent, then, at the level of the portfolio, the central limit theorem applies, and internal funds are approximately normally distributed.

internal funds make future investment opportunities more costly to finance because of deadweight financing costs, $C(e)$. Thus, negatively skewed payoffs will tend to increase the expected deadweight costs of external finance, thereby directly reducing value. This leads to a positive $\widetilde{G}$ and a reduced level of $P(w)$, all else equal.[22] Asymmetries here affect expected net cashflows, and thereby affect value.

While the continuous trading argument above might plausibly apply to a company's equity, it certainly wouldn't apply to the illiquid insurance exposures that comprise w. Asymmetries in the illiquid insurance exposures can reduce $P(w)$ in our model because, for these risks, a continuous trading argument is not tenable. There is essentially no trading in individual illiquid risks. This is what allows financial intermediaries and, indeed, most non-financial firms as well, to add value. They originate and warehouse illiquid assets and exposures, allowing them in the aggregate to be liquidly traded in the form of company stock.

Proposition 4 allows for distributional asymmetries in valuation, but not by including additional factors driven by skewness, or other higher-order moments, per se. The problem with a moment-based approach is that there are, of course, an infinite number of moments and the criteria for choosing among them can be rather arbitrary. The advantage of our functional approach is that all of the properties of the asymmetric distribution – in effect all of its moments – are contained in our single third factor.

However, there are also disadvantages to this analytic approach. Chief among them is that we cannot easily allow for non-normally distributed new opportunities. Our method for deriving explicit solutions – the generalization of Stein's Lemma – can be applied to at most one nonlinear function of normals at a time. Essentially, there is no general analytic solution for the covariance of two non-normal random variables. Our response is to assume that the new opportunities are distributed normally (and therefore symmetrically). This works, but leads to pricing errors when new opportunities are asymmetrically distributed.

---

[22] The distributional asymmetry could impact $P(w)$ through the product-market channel as well. For example, if the negative demand shocks associated with greater risk-to-capital ratios negatively affected the return on future investments, $F(I) = F(I; s^2)$, $\widetilde{G}$ would operate through the product-market as well as the capital-market channel. Indeed, an earlier version of this paper included such argument of F. It was dropped because, while probably accurate, it added more to the model's complexity than to its intuition.

Consider, for example, an insurer exposed to asymmetrically-distributed property-casualty losses. An excess-of-loss reinsurance contract (essentially a call spread written on the insurer's underlying event losses) would commonly be used to cede the risk. Such a contract would, of course, have a highly asymmetric payoff distribution. The contract would be *more* valuable to the insurer than a contract whose payoffs were symmetric (and normal) because the excess-of-loss contract would completely cede the insurer's event losses without adding any other risks, something that a contract with symmetric payoffs could not do. By approximating the new opportunity payoffs as normal, our approach therefore understates value if a positively asymmetric new opportunity perfectly hedges pre-existing exposures, and overstates value if there are no pre-existing exposures. Nevertheless, our contribution is to provide at least a basis for pricing new opportunities that alter the profile of pre-existing asymmetric exposures.

E. Simultaneous multiple new opportunities

Thus far, in solving for the optimal quantity and hurdle rate for the jth new opportunity, we have taken as given all the other J quantities and hurdle rates. It is as though the insurer is considering a single new opportunity, having already fixed decisions on the J-1 others. Clearly, the optimal thing to do is to solve the J first-order equations simultaneously. This is straightforward to do. To express the results, we use notation in which n^* is a Jx1 vector of optimal amounts of new opportunities; Σ is the JxJ covariance matrix of the shocks of the unhedgeable insurance components, $\varepsilon_{N,j}^I$; μ_N is the Jx1 vector of total expected returns on the new opportunities; C_{NM} is the Jx1 vector of covariances of the new opportunities with the market, $\text{cov}(\varepsilon_{N,j}, M)$; and C_{NP} is the Jx1 vector of covariances of the new opportunities with the pre-existing exposures, $\text{cov}(\varepsilon_{N,j}, \varepsilon_P)$. The vector of optimal investments in new opportunities is given by:

$$n^* = \left(\frac{1}{F+G}\right)\Sigma^{-1}\left(\mu_N - \gamma C_{NM}\right) - \left(1 + \frac{\tilde{G}}{F+G}\right)\Sigma^{-1}C_{NP},$$

(14)

whereas the vector of incremental required returns for an incremental opportunity is:

$$\mu_N = \gamma C_{NM} + (F + G)(C_{NP} + \Sigma n) + \widetilde{G} C_{NP}, \tag{15}$$

Both of these expressions are solutions to the J first-order conditions in Propositions 3 and 4. However, simultaneous solution makes the coordination problem among the J individual decisions clear. The decision to hold more or less the jth opportunity impacts the desirability of the other J-1 opportunities. If the decisions must be made simultaneously, a central planner would need to implement equation (14). The planner would need to collect all information on the J opportunities and then mandate the actions. Of course, in practice, such micro-coordination is rarely even a goal, let alone an actual achievement.

One reason that organizations do not attempt such coordination is that the individual decisions are typically small relative to the firm-wide capital. In that case, coordination is less important. We can approximate this above by treating each of the opportunities as small. In that case, the hurdle rate in Proposition 4 can be applied independently to each opportunity. That would imply that the incremental hurdle rate for an infinitesimal amount of the J new opportunities is:[23]

$$\mu_N = \gamma C_{NM} + (F + G + \widetilde{G}) C_{NP}. \tag{16}$$

This expression further highlights the reason that we refer to the hurdle rates as 'incremental.' Usually, hurdle rates are independent both of the amount being invested and of other investments. That is not the case here. The jth hurdle rate depends not only on how much investment takes place in the other J-1 opportunities; it depends also on n_j itself. The hurdle rate

---

[23] Equation (16) has two distinct factors, rather than the three in Proposition 4. However, that turns out to be a special feature of our simplifying assumptions and will not obtain generally. If we were to divide additively the asymmetrically-distributed disturbance, ε_P^I, into asymmetric and normally-distributed components, then three factors would be required to express the hurdle rate.

in Proposition 4 reflects this directly through the second priced factor, $\text{cov}(w, \varepsilon_{N,j}^I) = \text{cov}(\varepsilon_P^I + \sum_i n_i \varepsilon_{N,i}^I, \varepsilon_{N,j}^I)$, which is a linear function of all J of the n's and an increasing linear function of n_j.

There are also indirect effects of n_j on the hurdle rate. As n_j increases, the factor loadings, F, G, and $\widetilde{G}$ will generally change as well. Thus, for example, if the insurer takes a large, very risky position in one opportunity – even if this position is independently distributed of all others – this will generally increase all the factor loadings, and thereby make the firm less willing to take on other risks. This is what Froot and Stein (1998) call a "firm-wide" risk effect: the distribution of internal funds, w, and therefore the factor loadings, are affected by changes in n_j.

Because $\mu_{N,j}$ is dependent on n_j, the hurdle rate is applicable only to an incremental new opportunity, where the contemplated opportunity is small and where n_j is the amount already in place. That is what we mean by an "incremental hurdle rate for an incremental new opportunity." For a discrete opportunity of size $\overline{n}_j$, one would need to integrate the expression in Proposition 4 over n_j to get the hurdle appropriate to a discrete-sized opportunity.

IV. Optimal Capital Structure at Time 0

We are now in a position to step back to time 0 and solve for the optimal capital level K. There is a simple tradeoff at work: on the one hand, as noted above, a higher K moves the components of the insurer's effective risk aversion, F, G, and $\widetilde{G}$ toward zero. Propositions 3 and 4 show that, ex ante, the insurer can invest more aggressively in new opportunities that promise an above-market return at time 1, to hedge less with products that appear very costly at time 1, and to cut back on overpaying for and/or overhedging with products whose chief purpose is to improve distributional asymmetries. All of these benefits, of course, must be balanced against the higher deadweight costs, τK, that come with higher levels of capital.

To illustrate the first part of the tradeoff most transparently, suppose that the new opportunity in question is a small one and that the other J-1 decisions are fixed. In this setting, a

natural question to ask is how the insurer's hurdle rate – as given by Proposition 4 -- changes with K:

$$\frac{d\mu_{N,j}}{dK} = \left(F_K + G_K\right)\text{cov}\left(w, \varepsilon_{N,j}^I\right) + \widetilde{G}_K \, \text{cov}\left(\varepsilon_P^I, \varepsilon_{N,j}^I\right) \tag{17}$$

The first two factor loadings, F_K and G_K, are unambiguously negative. If the covariance of the new opportunity with the pre-existing portfolio is positive, then this factor will push the hurdle rate smoothly toward fair market value as the amount of capital K is increased. It is a bit more complex for $\widetilde{G}_K$, which will be negative (positive) for negatively (positively) asymmetric risks. Generally, $\widetilde{G}$ will move toward zero from whatever side of zero it is on. Thus, any difference between firms with positive and negative $\widetilde{G}$'s will be mitigated by more capital. We pick up on the cross sectional implications of this in section VI below.

As of time 0, the insurer's objective function is to pick K so as to maximize $V - K$, recognizing that $V = V\left(w(\mu_P(n,K), n(K), K)\right)$. In words, K affects w directly through the amount of financial slack that will be available at time 2, as well as indirectly through its influence on the optimal choice of new opportunities, n, and through its influence on the product market expected opportunity set, μ_P. Fortunately, one can use the envelope theorem to show that the solution to this problem can be written simply as:

$$\tau = F\frac{\sigma^2}{K} + 1 - \left(\frac{1+r}{E[P_w]}\right). \tag{18}$$

Equation (18) has an intuitive interpretation. The insurer can hold another dollar of slack capital at time 0 and pay the costs, τ, of carrying that capital. By spending the additional deadweight cost, the firm earns two benefits. First, it benefits from the improved product-market terms of trade. This is summarized by $F\frac{\sigma^2}{K} > 0$ – the marginal impact on product market terms of trade of additional capital, holding risk constant. Second, it benefits from the reduction in the

expected costs of external funds. This is summarized by $\dfrac{E[P_w]-(1+r)}{E[P_w]} > 0$, which is literally the

percentage gain from substituting for an expensive dollar of external capital the (now cheap) future value of an additional dollar of internal capital. Clearly, the optimal level of K will equate marginal costs against marginal benefits.

In the limiting case where $\tau = 0$, so that there are no deadweight costs of holding capital, the insurer holds an arbitrarily large amount. As K then becomes large, the expected shadow value of external funds, $E[P_w]$, falls to one. This in turn implies that both G and F converge to zero as well. The insurer then behaves in a classical manner, making pricing and allocation decisions according to a purely market-based model of risk and return. In contrast, as τ increases above 0, the insurer holds less capital, thereby raising its effective risk aversion, and amplifying the deviations from textbook capital budgeting principles.[24]

Notice also that the existence of product-market imperfections encourages the insurer to hold more capital than would be the case if there were only capital-market imperfections. These imperfections make the insurer more risk averse, increasing its demands for hedging and capital, and reducing its appetite to bear risk through asset holdings or underwriting.

V. Some Implications of the Three-Factor Model

Adding the third factor – asymmetry – provides another channel through which cross sectional variation in firms' appetite for risk management might be explained. Leaving aside for the moment the product market effects summarized by F, the asymmetry factor suggests that the costs of external finance, as summarized in the $P(w)$ function, will not be sufficient to explain optimal firm behavior.

Consider, for example, large "value" firms, which are most likely to have negatively asymmetric distributions – they have a lot to lose, but only relatively smaller amounts to gain. At the same time, they have relatively solid access to external markets for capital. A good example

---

[24] While G and $\widetilde{G}$ do not appear explicitly in equation (18), both the concavity of P and the degree of negative asymmetry of distribution of w have, all else equal, negative impacts on $E[P(w)]$.

might be Philip Morris, which has high market share in its product markets, large potential legal liabilities, and is one of the largest firms by profits in the world. On the basis of the two-factor model alone, one might conclude that Philip Morris with its relatively free access to the capital market might not optimally impose important deviations from fair-market valuation in assessing new opportunities. However, the third factor suggests that concavity of P comes more from the distribution of outcomes of w and less from the curvature of the function itself. As a result, the firm may optimally have relatively higher hurdle rates than might be expected on the basis of the first two factors alone.

Small "growth" firms, by contrast, behave in just the opposite way. They are most likely to have positively asymmetric distributions of their assets (including intangibles). Partially offsetting this, they may have very concave P functions because of their small size and difficulty in reaching the capital markets. For these firms, the asymmetry factor may tend to reduce hurdle rates from what would appear from the two-factor model as high levels, thereby mitigating the expected deadweight costs of external finance.

It is important to note in this latter case that these reductions in hurdle rates are limited. That is, the third factor can never be so large as to overtake the second factor. In other words, the sum $G + \widetilde{G}$ is always positive. Thus, regardless of the sign of $\widetilde{G}$, the second and third factors together unambiguously raise hurdle rates for positively correlated opportunities.

The asymmetry factor also seems to apply naturally to the insurance and reinsurance sectors. As mentioned above, insurers and reinsurers, as aggregators of catastrophic exposures, tend to have negatively asymmetric payoff distributions. Unlike other financial institutions, such as commercial or savings banks, insurance companies do not have a capital provider of "last resort" to mitigate the impact of the negative asymmetry. These concerns may weigh heavily on customers and capital providers who are valuing the firm.

As mentioned above, it would be desirable to extend the results to allow for asymmetries in the distribution of new opportunity outcomes as well. Unfortunately, we cannot solve explicitly for non-normal new opportunity disturbances, and we do not pursue here this more general case. However, we do not need restrictive assumptions about *any* distributions in order

to reach equation (9), prior to the application of the generalization of Stein's Lemma.[25] Thus, while we cannot solve explicitly for n_j or $\mu_{N,j}$ if the jth new opportunity is non-normally distributed, the first-order condition in (9) always applies.

Equation (9) emphasizes that the incremental value of an opportunity is a linear function of the covariance between the distributions of the jth opportunity and P_w, the shadow value of time-2 internal funds. All else equal, the larger the covariance, the greater the marginal impact on firm-wide risk, and, therefore, the greater the deviation from fair value will tend to be. Thus, from the firm's perspective, hedges are more valuable if they can eliminate greater amounts of variation in P_w, and new investments are more costly if they contribute more to variation in P_w. The P_w covariance term in equation (9) summarizes that effect.

Consider, for example, an opportunity such as a reinsurance contract that pays only under very high losses (and therefore is triggered with low probability). Such a contract is most likely to have large covariance with P_w. Most insurers or reinsurers hold a considerable amount of catastrophic event risk based on their portfolio of underwriting contracts. Because P_w is large when w is very low, and because w is likely to be very low because of a catastrophe, an asymmetric hedge that pays off if and only if a catastrophe occurs can be very valuable to the firm in terms of reduced variation in P_w. In spite of the small probability of a large loss, the negative asymmetry of $\varepsilon_{N,j}^{I}$ can make the P_w covariance term in equation (9) absolutely larger, since the asymmetry concentrates contract payoffs into high-P_w states of nature.

Perhaps it is not so surprising then, that insurers and reinsurers, who warehouse catastrophic exposures, find it very expensive to hedge them, at least among one another. In fact, there is evidence that realized prices of reinsurance tend to become much more expensive relative to fair value as contract retentions (i.e., deductibles) increase.[26] Taken literally, this suggests that more negatively asymmetric exposures are most costly for writers to take on and most valuable for cedents to shed.

---

[25] Except that we do need to assume that second own-moment and cross-moments exist.
[26] See Froot (1996).

VI.   Asymmetrically Distributed New Opportunities

This last section modifies an approach, discussed in Froot and Stein (1998), for valuing new opportunities that are asymmetrically distributed. This approach gives some useful insight; however, it can be applied in the case of insurance exposures only in rather limited circumstances, which we clarify below.   To fix ideas, imagine that a reinsurer has a new opportunity to write a very small amount of excess-of-loss cover for the catastrophe losses of an insurer's underwriting book. (The excess-of-loss contract is similar to a call spread using standard options.) Imagine also that the insurer has issued a capital market instrument with payoffs tied directly linearly to its time 2 catastrophe results.[27]

We need to assume that the price of this underlying instrument, *e,* follows a simple stochastic process between times 1 and 2, i.e., geometric Browian motion with drift θ and instantaneous standard deviation, υ:

$$de = \theta e + \upsilon e dz \ .$$
(19)

From Proposition 4, we can define the excess return that the insurer requires on the underlying position as [28]

$$\mu_{N,e} = \gamma \, \text{cov}\left(M, \frac{de}{e} \right) + (F+G) \text{cov}\left(w, \frac{de}{e} \right) + \tilde{G} \, \text{cov}\left(\varepsilon_P^I, \frac{de}{e} \right)$$
(20)

By Ito's lemma, the instantaneous change in the value of the derivative, *L*, is given by:

---

[27] Such an instrument is not really the same as what are usually called 'cat bonds,' in that it would not have a maximum or promised payment of interest and principal, but rather payments that are directly linked to the catastrophe component of the firm's underwriting. It is closer to a market traded quota-share contract than to a catastrophe bond.

[28] In this continuous-time setting, ε_P^I should now be interpreted as the instantaneous innovation in the rate of return on the insurance component of the pre-existing portfolio.

$$dL = \left(\theta e L_e + L_t + \frac{\upsilon^2 e^2}{2} L_{ee} \right) dt + \upsilon e L_e dz \qquad (21)$$

That is, the derivative and the underlying have perfectly correlated instantaneously normal innovations, implying that $\text{cov}(dL, x) = \text{cov}(de L_e, x) = L_e \, \text{cov}(de, x)$. Applying Proposition 4 then yields the reinsurer's required excess return on the derivative:

$$\mu_{N,l} = \text{cov}\left(M, \frac{dL}{L} \right) + (F + G)\text{cov}\left(w, \frac{dL}{L} \right) + \widetilde{G}\,\text{cov}\left(\varepsilon_P^I, \frac{dL}{L} \right) \qquad (22)$$

This is the "trick" in applying our pricing equation to the asymmetric new opportunity, L. The continuous-time formulation implies that instantaneous innovations in the price of the derivative, dL, are normal, and this allows us to write equation (22). Over discrete intervals, the distribution of ΔL is not normal because the expected return and standard deviation of the dL process in equation (21) are changing over time.

The linearity of the covariance operator allows equation (22) to be rewritten as:

$$\mu_{N,l} = \frac{e L_e}{L}\left(\gamma\,\text{cov}\left(M, \frac{de}{e} \right) + (F + G)\text{cov}\left(w, \frac{de}{e} \right) + \widetilde{G}\,\text{cov}\left(\varepsilon_P^I, \frac{de}{e} \right) \right)$$
$$= \frac{e L_e}{L}\left(\mu_{N,e} \right) \qquad (23)$$

This says simply that the expected return required by the reinsurer on the derivative is equal to the excess return required on the underlying times the elasticity of the price of the derivative with respect to the price of the underlying, $\frac{e L_e}{L}$. In the case of standard call options, for example, this elasticity is always greater than one. Moreover, the elasticity is increasing in the strike price of the call, e.g., more out-of-the-money options have greater price elasticities with respect to the price of the underlying. The firm will charge a higher required return for a call option the higher is its strike

price. In the context of our example, this implies that the reinsurer will require a higher return on the excess-of-loss contract than on the underlying. In addition, this required return is more sensitive than the required return on the underlying with respect to F, G, and $\left|\widetilde{G}\right|$.

The advantages of this approach are several. First, it allows us to price the reinsurance contract, or, for that matter, any derivative written not only on the underlying cat book, but on *any* underlying illiquid instrument. The reinsurance contract will have payoffs that are far more negatively asymmetric than those of the underlying, all the more so as the layer is higher (i.e., as the retention or deductible on the reinsurance contract grows).

One can see that these results are useful for providing a flavor for the impacts of greater asymmetry. Higher reinsurance layers are like more out of the money call options, i.e., the probability of payoff is lower, and, therefore, the return is driven more by extreme outcomes. As a result a higher proportionate return is required – all three pricing factors are blown up by the elasticity, which is greater than one. Loosely speaking, the greater is the negative asymmetry of the new opportunity, the greater is the reinsurer's required return. Notice that the absolute increase in required return is an increasing function of $\widetilde{G}$, so that the importance of this effect is sensitive to the pre-existing asymmetry in the insurer's payoffs.

We should, however, be careful not to overstate the usefulness of these results. There are important assumptions, which, if not met would invalidate this analysis. The first is that the underlying cat instrument be continuously priced in the marketplace. For example, it would not be enough to observe the continuous time payoffs on the underlying cat book itself – we would need a price for these payoffs as determined in a relatively efficient marketplace.

The second is that if the underlying were costlessly tradeable, it could be readily priced and delta-hedged, and consequently the reinsurer would be able to lay off the risk without friction. Presumably, it would do this if the price were fair. So the underlying must also be illiquid.

The third restriction is that we assumed the reinsurance contract to be very small. This is critical because, as a cat loss emerges and e falls over time, a larger position would have resulted in a matching reduction in w, which would have increased F, G, and $\left|\widetilde{G}\right|$, and, in turn, increased

the hurdle rate over time. For the derivation above, we assumed that the expected returns are constant.

Finally, the continuous time formulation allows for us to write the new investment opportunity as continuously normal, and therefore apply our pricing formulas. The Faustian bargain for doing this, however, is that the required excess return on the derivative changes constantly through time as the elasticity, $\dfrac{eL_e}{L}$, evolves. So there is no single answer for the reinsurer's required return between times 1 and 2 as we had for other instruments.

VII.   Conclusions

In trying to better adapt internal pricing models to insurers and reinsurers, we have added two critical components to more standard capital markets imperfections: customer sensitivity to the level of insurer risk and asymmetrically distributed risk exposures.   Both of these components contribute directly to a value-based rationale for corporate risk management and prudent risk allocation. The 'excess' sensitivity of customers to insurer risk provide another channel, separate from the capital markets imperfections, through which the impact on firm-wide risk matters for pricing and allocation. Negatively asymmetric distributions of insurer-held risk contribute to insurer risk aversion through both the capital markets and product markets channels.

One question that inevitably arises is whether the model results should be interpreted as positive or normative. Certainly, they are positive in the sense that insurers and reinsurers *are*, in practice, concerned with risk management and capital allocation. Managers perceive that the Modigliani-Miller irrelevance theorems fail and, therefore, see that risk management can raise value. The results are positive in the sense that they provide a possible explanation for why reinsurance prices are so high (relative to expected loss), for why reinsurance prices for particularly large-scale catastrophic perils, like Florida wind, are among the highest, and for why prices rise and quantities of reinsurance supplied fall in the aftermath of large event losses.

However, there are clearly normative aspects to the theory as well.   Practitioners may

use cost of capital formulations that are derived in ad hoc ways and that aren't necessarily rooted in corporate value maximization. The theory may, therefore, have prescriptive content. Yet the precise results are also sensitive to the nature of the capital and product market distortions. So the overall approach may point in the right direction even if the exact results cannot be taken too literally.

VIII.   <u>References</u>

Cagle, J., and Scott Harrington, 1995, "Endogenous Supply with Capacity Constraints and Endogenous Insolvency Risk," *Journal of Risk and Uncertainty*, 11, 219-232.

Christofides, Stavros, and Andrew Smith, 2002, "DFA – the value of risk," unpublished manuscript, schristofides@bwdeloitte.com.

Cummins, J. David, and P.M. Danzon, 1997, "Price, Financial Quality, and Capital Flows in Insurance Markets," *Journal of Financial Intermediation,* 6, 3-38.

Cummins, J. David, and R.D. Phillips, 2000, "Applications of Financial Pricing Models in Property Liability Insurance," in *The Handbook of Insurance Economics,* G. Dionnes, ed., Boston: Kluwer Academic Publishers.

Cummins, J. David, Richard D. Phillips, and Stephen D. Smith, 1998, "Derivatives and Corporate Risk Management: Participation and Volume Decisions in the Insurance Industry," Working paper 98-19, Financial Institutions Center, Wharton School.

Cummins, J. David, and D.W. Sommer, 1996, "Capital and Risk in Property-Liability Insurance Markets," *Journal of Banking and Finance,* 20, 1069-1092.

Doherty, Neil A., 1991, "The Design of Insurance Contracts When Liability Rules are Unstable," *Journal of Risk and Insurance*, 61, 691-706.

Doherty, Neil A., and S.M. Tinic, 1982, "A Note on Reinsurance Under Conditions of Capital Market Equilibrium," *Journal of Finance*, 36, 949-953.

Epermanis, Karen, and Scott E. Harrington, 2001, "Market Discipline and Reaction to Rating Changes in U.S. Property-Liability Insurance Markets," presented at NBER Insurance Project Conference, February.

Froot, Kenneth, A., 1999, "The Limited Financing of Catastrophe Risk: An Overview," in K. Froot, ed., *The Financing of Catastrophe Risk,* University of Chicago Press, 1-22.

Froot, Kenneth, A., 2001, "The Market for Catastrophe Risk: A Clinical Examination," *Journal of Financial Economics*, 60, 529-571.

Froot, Kenneth A. and Paul G.J. O'Connell, 1997, "On the Pricing of Intermediated Risks: Theory and Application to Catastrophe Reinsurance," Harvard Business School working paper.

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein, 1993, "Risk Management: Coordinating Corporate Investment and Financing Policies," *Journal of Finance,* 48, 1629-1658.

Froot, Kenneth A., and Jeremy C. Stein, 1998,"Risk Management, Capital Budgeting and Capital Structure Policy for Financial Institutions: An Integrated Approach," *Journal of Financial Economics*, 47, 55-82.

Gale, Douglas, and Martin Hellwig, 1985, "Incentive-Compatible Debt Contracts I: The One-Period Problem," *Review of Economic Studies,* 52, 647-664.

Garven, James, R., and Joan Lamm-Tennant, 1999, "Economic and Financial Perspectives on the Demand for Reinsurance," working paper, Baylor University.

Grace, Martin F., Robert W. Klein, and Paul R. Kleindorfer, 2001, "The Demand for Homeowners Insurance with Bundled Catastrophe Coverage," Presented at NBER Insurance Project Conference, February.

Grace, Martin F., Robert W. Klein, Paul R. Kleindorfer, and Michael R. Murray, 2003, "Catastrophe Insurance: Consumer Demand, Markets, and Regulation," Boston: Kluwer Academic Publishers.

Greenwald, Bruce C., Alec Levinson, and Joseph E. Stiglitz, 1991, "Capital Market Imperfections and Regional Economic Development," in Giovannini, ed.: *Finance and Development: Lessons and Experience* CEPR, 65-93.

Gron, Anne, 1994a, "Capacity Constraints and Cycles in Property-Casualty Insurance Markets," *Rand Journal of Economics,* 25, 110-127.

Gron, Anne, 1994b, "Evidence of Capacity Constraints in Insurance Markets," *Journal of Law and Economics*, 25, 110-127.

Hall, Brian, 1999, "Property and Casualty Solvency Funds as a Tax and Social Insurance System," in K. Froot, ed., *The Financing of Catastrophe Risk*, University of Chicago Press.

Hoerger, T.J., F.A. Sloan, and M. Hassan, 1990, "Loss Volatility, Bankruptcy, and the Demand for Reinsurance," *Journal of Risk and Uncertainty,* 3, 221-245.

Kahneman, Daniel, and Amos Tversky, 1979, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, Vol. 47, No. 2, March, 263-291.

Kozik, Thomas J., and Aaron M. Larson, 2001, "The n-Moment Insurance CAPM, *Proceedings of the Casualty Actuarial Society*, LXXXVII, 168-169 (May).

Kraus, Alan, and Robert Litzenberger, 1976, "Skewness Preference and the Valuation of Risky Assets," *Journal of Finance,* 1085-1100.

Merton, Robert, C., 1993, "Operation and Regulation in Financial Intermediation: A Functional Perspective," in P. Englund (ed.), *Operation and Regulation of Financial Markets*, 17–68, The Economic Council, Stockholm.

Merton, Robert C., 1995a, "A Model of Contract Guarantees for Credit-Sensitive Opaque Financial Intermediaries," Harvard Business School working paper.

Merton, Robert C., 1995b, "A Functional Perspective on Financial Intermediation," *Financial Management,* 24, 23-41.

Merton, Robert C., and André F. Perold, 1993, "The Theory of Risk Capital in Financial Firms," *Journal of Applied Corporate Finance,* 5, 16-32.

Myers, Stewart C., 1984, "The Capital Structure Puzzle," *Journal of Finance,* 39, 575-592.

Myers, Stewart C., and Nicholas Majluf, 1984, "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have," *Journal of Financial Economics,* 13, 187-221.

Myers, Stewart C. and J. A. Read, 2001, "Capital Allocation for Insurance Companies," *Journal of Risk and Insurance*, 68, 545-580.

Phillips, Richard D., J. David Cummins, and Franklin Allen, 1998, "Financial Pricing of Insurance in Multiple Line Insurance Companies," *Journal of Risk and Insurance*, 65, 597-636.

Raiffa, Howard, 1969, "Preferences for Multi-Attributed Alternatives," Memorandum RM-5868-DOT/RC (Santa Monica: Rand Corporation, April), 81-95.

Sharpe, Steven A., 1995, "Bank Capitalization, Regulation, and the Credit Crunch: A Critical Review of the Research Findings," Federal Reserve Board working paper # 95-20.

Sommer, David, W., 1996, "The Impact of Firm Risk on Property-Liability Insurance Prices," *Journal of Risk and Insurance,* 63, (September), 501-514.

Stein, Charles, 1981, "Estimation of the Mean of a Multivariate Normal Distribution," The Annals of Statistics, 9, 1135-1151.

Stein, Jeremy C., 1998, "An Adverse Selection Model of Bank Asset and Liability Management with Implications for the Transmission of Monetary Policy," *RAND Journal of Economics,* 29, Autumn, pp. 466-486.

Taylor, G., 1994, "An Equilibrium Model of Insurance Pricing and Capitalization," *Journal of Risk and Insurance,* 62, 409-446.

Townsend, Robert M., 1979, "Optimal Contracts and Competitive Markets and Costly State Verification," *Journal of Economic Theory,* 21, 265-293.

Venter, Gary G., and John A. Major, "Allocating Capital by Risk Measures: A Systematic Survey," Guy Carpenter, 2002.

Wakker, Peter P., Richard H. Thaler, and Amos Tversky, 1997, "Probabilistic Insurance," *Journal of Risk and Uncertainty,* 15, 7-28.

Weinstein, Milton C., Donald S. Shepard, and Joseph S. Pliskin, 1980, "The Economic Value of Changing Mortality Probabilities: A Decision-Theoretic Approach," *Quarterly Journal of Economic*, Vol. 94, No. 2 (March), 373-396.

Weinstein, Milton C., and Robert J. Quinn, 1983, "Psychological Considerations in Valuing Health Risk Reductions," *Natural Resources Journal*, 23 (July), 659-673.

Winter, Ralph A., 1994, "The Dynamics of Competitive Insurance Markets," *Journal of Financial Intermediation,* 3, 379-415.

Zanjani, George, 2002, "Pricing and Capital Allocation in Catastrophe Insurance," *Journal of Financial Economics*, 65, 2 (August), 283-305.