Risk Management and Insurance Review

© *Risk Management and Insurance Review*, 2014, Vol. 17, No. 2, 137-161
DOI: 10.1111/rmir.12022

FEATURE ARTICLES

Optimal Insurance With Costly Internal Capital

Dwight Jaffee
Johan Walden

Abstract

We introduce costly internal capital into a standard insurance model, in which a risk-averse policyholder buys insurance from a risk-neutral insurer with limited liability. The unique optimal contract and internal capital lead to a strictly positive probability for insurer default. Some risks are uninsurable in that the insurer chooses not to provide insurance against such risks. An increase in the cost of capital may lead to a higher optimal amount of internal capital. The results extend to multiple policyholders in a symmetric setting. Our extension of the classical model to include costly internal capital provides a fruitful approach to many real world insurance markets.

Introduction

If insurers hold sufficient capital, they can make a credible guarantee to pay all claims. As shown in the seminal analysis by Arrow (1963),[1] the optimal insurance contract between a risk-neutral insurer and a risk-averse policyholder in this case is full coverage against losses above a strictly positive deductible. In practice, two factors together make such a contract infeasible. First, virtually all insurers are now limited liability corporations, which eliminates the unlimited recourse to partners' external (private) assets that was once common.[2] To avoid counterparty risk, a large amount of capital therefore needs to be held within the firm. Second, the excess costs of holding internal (on balance sheet) capital, such as corporate taxes, asymmetric information, and agency costs, provide a strong incentive for insurers to limit the amount of such capital they hold.

This article was subject to double-blind peer review.
Dwight Jaffee is at the Haas School of Business, University of California at Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720-1900; phone: 510-642-1273; e-mail: jaffee@haas.berkeley.edu.
Johan Walden is at the Haas School of Business, University of California at Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720-1900; phone: 510-643-0547; e-mail: walden@haas.berkeley.edu.
[1] See also Arrow (1971, 1974).
[2] The insurer Lloyds of London once provided a credible guarantee to pay all claims, based on the private wealth of its "names" partners. In the aftermath of large asbestos claims, however, Lloyds now operates primarily as a standard "reserve" insurer with balance sheet capital and limited liability.

The excess costs of holding internal capital, for example, in the form of lost tax shields, have long been studied in the corporate finance literature; see, for example, Modigliani and Miller (1963), and more recently Froot et al. (1993) and Froot and Stein (1998). For insurance firms, Cummins (1993), Merton and Perold (1993), Jaffee and Russell (1997), Myers and Read (2001), and Froot (2007) all emphasize the importance of various accounting, agency, informational, regulatory, and tax factors in raising the cost of internally held capital. Zanjani (2002) argues that the cost of internal capital may be especially important for catastrophic insurance. In practice, policyholders therefore face counterparty risk.

The risk of insurer default in paying policyholder claims has led to the imposition of strong regulatory constraints on the insurance industry in most countries. Capital requirements are one common form of regulation, although no systematic framework is available for determining the appropriate levels.[3] As Cummins (1993) and Myers and Read (2001) point out, it is likely that the capital requirements are being set too high in some jurisdictions and too low in others, and similarly for the various lines of insurance risk, in both cases leading to inefficiency. It is thus important to have an objective framework for identifying the appropriate level of capital based on each insurer's particular book of business. In this article, we provide such a framework.

We study the design and properties of an optimal contract in a competitive insurance market, between a risk-neutral insurer and a risk-averse policyholder, when the insurer has limited liability and internal capital is costly. The model endogenously determines the optimal level of the deductible, the policy premium, and the capital held by the insurer (which in turn determines the states in which the insurer defaults). To our knowledge, the optimal contract and its characteristics under these conditions of costly internal capital have not been previously studied, providing our first contribution. The optimal contract itself is shown—in line with previous literature—to be one with a deductible. Above the deductible, full insurance is available until all internal capital is used; above that point the insurer defaults. The optimal contract thus depends on two parameters: the deductible and the amount of internal capital. The insurance contract can be viewed as the difference between a call option for the policyholder on the loss with the strike price being the deductible, and a call option for the insurer to default, with the strike price being the sum of internal capital and the deductible. Using standard option pricing arguments, we characterize the premium charged for the insurance.

For any given (strictly positive) level of internal capital, the deductible is shown to be unique, and there is a unique level of internal capital that leads to a globally optimal contract. This contract is associated with a strictly positive probability of insurer default. As long as the cost of internal capital is lower than the amount of capital held, no risk is

[3] Insurance guaranty funds represent a second form of regulation that attempts to control the risk of insurer default. These funds are required in many states for consumer lines, principally homeowner and auto insurance. All of the insurers participating in the line are required to provide funds to meet policyholder claims in case another insurer has failed. While the guaranty funds may reduce the risk of insurer default by diversifying the risk, Cummins (1988) points out that even for the applicable lines, policyholders still face substantial default costs since the guaranty funds usually impose a maximum payment per claim and because the payments may be substantially delayed as the claims are transferred from the bankrupt insurer to the guaranty fund.

totally uninsurable in the sense that there will always be some policyholder preferences for which insurance demand is positive. The effects of an increase in the cost of internal capital on the optimal deductible and amount of internal capital are ambiguous. For the special case when losses have a two-point distribution, the analysis simplifies significantly. We provide a full characterization of the solution for this specific case.

The assumption of a single policyholder has been commonly used in the insurance literature, including Arrow (1963), but may seem restrictive in this setting. However, the assumption may best be viewed as the policyholder being a "representative agent" for multiple policyholders, a common approach used in both the finance and insurance literatures. This interpretation is especially straightforward if the multiple policyholders face highly correlated risks—effectively the *same* risk. In this case, the single policyholder model can be directly reinterpreted as representing the (aligned) utilities of all agents. Insurance against natural disasters, like earthquakes, provides an example where such an interpretation is fairly straightforward. More generally, we show that the representative agent interpretation remains valid in situations where policyholders are "symmetric" in their preferences and the risks they face. Thus, our single-policyholder model is not restrictive in a variety of relevant cases. This extension to multiple policyholders provides our second contribution. The completely general case, with asymmetries in preferences and/or risk structures, is outside of the scope of our model and remains an interesting topic for future research.

Our model and results extend the ongoing research that began with the classical model in Arrow (1963). Schlesinger (1981) allows for general nonlinear excess cost functions for the insurer, as a function of expected losses. Eeckhoudt and Gollier (1999) study the case when the policyholder faces multiple risks. Meyer and Ormiston (1999) introduce a method to analyze whether the optimal deductible is unique. None of these studies consider the case with limited liability, which is the focus of our study. Doherty and Schlesinger (1990) analyze the optimal insurance in case of possible insurer default, but default is modeled as an exogenous event in their paper, as is the case in Cummins and Mahul (2003) and Mahul and Wright (2004, 2007). Cummins and Mahul (2004) allow for an upper bound on compensation, for example, because of limited liability, but treat this bound as exogenous. Huberman et al. (1983) study optimal contracts when the *policyholder* has limited liability. Their setting is thus different from ours.

Our analysis is most closely related to Cummins and Mahul (2004), with our contribution being that internal capital is costly in our model and that we solve for the optimal amount of capital for the insurer to hold in this case. Kaluszka and Okolewski (2008) allow for an insurance premium that depends on both expected and maximal claims, but do not cover our case where costly internal capital provides the sole deviation of the premium from expected claims. Doherty et al. (2011) show that when losses are nonverifiable, upper bounds on compensation naturally arise. Thus, their optimal contract is similar to ours, but their mechanism is quite different. Importantly, another difference is that all of the above papers focus on the case with a single policyholder, whereas we extend our results to multiple policyholders in a symmetric setting.

The article is organized as follows. In the second section we introduce an insurance model with costly internal capital and analyze the optimal contract and premium given an exogenously specified level of internal capital. In the third section we endogenize the level of internal capital, and in the fourth section we discuss discrete loss distributions, specifically focusing on the two-point distribution. In the fifth section we extend the main results multiple policyholders in a symmetric setting. The final section concludes.

A MODEL WITH COSTLY INTERNAL CAPITAL

At $t = 0$, an *insurer* (i.e., an insurance company) in a competitive insurance market sells insurance against an observable idiosyncratic risk, $L \geq 0$ (throughout the article we use the convention that losses take on positive values) to an *insuree*. We assume that L has an absolutely continuous distribution, with a strictly positive probability density function, $\phi(x)$, on the positive real axis.[4] We also define the complementary cumulative distribution function $\Phi(x) = \int_x^\infty \varphi(y)dy$. We use the notation $\mu_X = E[X]$ for the expectation of a general random variable, X.

The risk is realized at $t = 1$, at which point the insurer makes a payment of $I(L)$ to the insuree. We call I the compensation (indemnity) function. The expected loss of the risk is $\mu_L = E[L] < 8$. The one-period discount rate is normalized to 0.

The insurer, who has limited liability, reserves capital of A within the company, and the maximal payment it can make is therefore bounded by A. Furthermore, in line with the literature, we assume that the compensation must be nonnegative and not greater than L. Thus, the constraint on the compensation function is:

$$0 \leq I(L) \leq \min(L, A). \tag{1}$$

Holding internal capital is costly. We assume that there is a proportional cost, so that the cost of holding A is δA, $0 < \delta < 1$. This is in line with a tax shield interpretation of what is the source of costly internal capital. Several of our results can be generalized to weakly convex cost functions, in line with the assumptions in Froot (2007). The insurer is risk neutral and the market is competitive, so the premium, P, paid by the insuree for insurance is

$$P = E[I(L)] + \delta A. \tag{2}$$

The insuree has a twice continuously differentiable expected utility function, u, defined on the negative real axis, such that $u' > 0$ and $u'' < 0$. Furthermore, we assume that $|Eu(-L)| < \infty$, that is, that the insuree's expected utility of taking on the whole risk is finite, as well as the technical conditions $Eu'(-L) < \infty$, and $|Eu''(-L)| < \infty$.

Because of the competitive market, the insurer aims at offering an optimal insurance contract leading to the following optimization problem:

$$\max_{I(L), A} E\left[u(-L + I(L) - P)\right], \quad \text{s.t. (1) and (2).} \tag{3}$$

As noted, this problem formulation has similarities to that in Cummins and Mahul (2004), but differs by introducing costly internal capital and by endogenizing A. The same is true

[4] Extensions to more general distributions are discussed further on in the article.

for the formulation in Kaluszka and Okolewski (2008), who consider pricing functions of the form $P = (1 - \delta)E[I(L)] + \delta A$, $0 < \delta < 1$. However, in our setting there is no δ in the first term on the right-hand side of (2).

The following result shows that, given internal capital of A, the optimal contract is a standard insurance contract with a deductible, above which all residual risk is insured until the limited liability constraint is reached.

Proposition 1: *The optimal insurance contract, given internal capital of $A > 0$, is*

$$\begin{aligned} I(L) &= 0, && L \leq D, \\ I(L) &= L - D, && D < L \geq D + A, \\ I(L) &= A, && L > D + A, \end{aligned}$$

for some unique deductible $D > 0$, which depends continuously on A.

In case $L > D + A$, the insurer is said to default. The optimal contract can thus be viewed as a "stop-loss" contract with a deductible.

This result is similar to what is shown in Cummins and Mahul (2004), the only difference being our assumption about costly internal capital. We could have taken their approach of proving it, by using the method in Meyer and Ormiston (1999) of reformulating the problem in terms of expected compensation. Instead, we prove the result (see the Appendix), using the direct problem formulation since this makes the mechanism behind the result more transparent.

The optimal insurance contract is thus characterized by a deductible D and the amount of internal capital, A. Since D is unique for each choice of internal capital, $A > 0$, we may also view D as a function of A and write $D(A)$. The first-order condition (FOC) for D will be analyzed in the next section—and is fairly complicated. For low levels of internal capital A, however, the characterization of the optimal D is simple. We define $\mu_{u'} = E[u'(-L)]$, and it then follows that:

Proposition 2: *As A tends to 0, the optimal deductible, D tends to Δ, defined as the solution to*

$$u'(-\Delta) = \mu_{u'}. \tag{4}$$

Thus, for low levels of internal capital, the deducible is chosen such that the marginal utility of a loss equal to the deductible is basically equal to the expected marginal utility without insurance, $\mu_{u'}$. The left-hand side of (4) represents the increase in marginal utility of decreasing the deductible, per unit of internal capital. The right-hand side of (4) represents the offsetting decrease of marginal utility from a higher premium when the deductible is decreased. At the optimum, Δ, the two effects are equal. We note that since u' is strictly decreasing, there is a unique solution to (4).

Technically, the optimal contract is degenerate when $A^* = 0$, since any D leads to the same outcome—namely, that the risk is uninsured. We disregard this degeneracy of the contract at $A = 0$, and use the convenient (as we shall see) convention that $A = 0, D = \Delta$ is *the* contract that corresponds to no insurance. This convention leads to continuity of $D(A)$ at $A = 0$.

The optimal contract can be viewed as the difference between two call options, one call option by the insuree to get payments net the deductible, $W(D) = \max(L - D, 0)$ and one call option by the insurer to default, $Q(A) = \max(L - D - A, 0)$, so that in total $I(L) = W(D) - Q(A)$. The premium can therefore be written

$$P = P(A, D) = \mu_W - \mu_Q + \delta A.$$

The following behavior of the premium as a function of D and A then follows immediately from standard option pricing theory:

Proposition 3: *The insurance premium, P(A,D) satisfies the following conditions:*

1. $P(0, D) = 0,$
2. $lim_{A\to\infty} P(A, D) - \delta A = \mu_W,\ for\ all\ D,$
3. $lim_{D\to\infty} P(A, D) = \delta A,\ for\ all\ A,$
4. $\frac{dP}{dA} = \delta + \Phi(A + D) > 0,$
5. $\frac{dP}{dD} = -\Phi(D) + \Phi(D + A) < 0,$
6. $\frac{d^2P}{dA^2} = \frac{d^2P}{dDdA} = -\varphi(A + D) < 0,$
7. $\frac{d^2P}{dD^2} = \varphi(D) - \varphi(D + A) \lesseqgtr 0.$

Under the additional condition that the probability density function, ϕ, is strictly decreasing, the sign of d^2P/dD^2 is unambiguous, $d^2P/dD^2 > 0$. Thus, the premium is well characterized.

OPTIMAL INTERNAL CAPITAL

Given the form of the optimal insurance contract in Proposition 1, and defining $F = Eu(-L + I(L) - P)$, it follows that the FOCs in (1) are given by

$$0 = \frac{\partial F}{\partial A} = \alpha - (\delta + \Phi(D + A))(\alpha + Zu'(-D - P) + \beta), \tag{5}$$

$$0 = \frac{\partial F}{\partial D} = Z\alpha - Z(1 - Z)u'(-D - P) + Z\beta, \tag{6}$$

where

$$Z = \Phi(D) - \Phi(D + A),$$

$$\alpha = \int_{D+A}^{\infty} u'(-x + A - P)\,\varphi(x)dx,$$

$$\beta = \int_{0}^{D} u'(-x - P)\varphi(x)dx.$$

The FOC with respect to D, (6), is necessary for an interior solution, and from Proposition 1 it is sufficient.

By rearranging the terms in (6), we arrive at

$$u'(-D - P) = \alpha + Zu'(-D - P) + \beta, \tag{7}$$

as the FOC in D for any given A. When plugged into (5), this leads to

$$\alpha = (\delta + \Phi(D + A))u'(-D - P), \tag{8}$$

as a necessary condition for a globally optimal interior solution. Condition (8) relates the insuree's marginal utility at the point where the insurance contracts begins to pay, $-D-P$, to the average marginal utility in states of default, adjusted for the extra costs of internal capital, $(\delta + \Phi(D + A))^{-1}\alpha = (\delta + \Phi(D + A))^{-1} \int_{D+A}^{\infty} u'(-x + A - P)\varphi(x)dx$. For a contract to be optimal they need to be equal, because otherwise the insuree can be made better off by changing A and D. We denote the optimal value(s) of the internal capital and deductible by A^* and D^*, respectively.

From the facts that uninsured expected utility is finite, $|Eu(-L)| < \infty$, and that the premium grows without bounds for large A (see (2)), it follows that $A^* = \infty$ cannot be a formal solution, but given the nonconcave behavior of the compensation function, I, and the premium, P, we would *a priori* not expect the optimal insurance contract to necessarily be either unique, nor interior. In general, we would expect solutions on the form (i) $A^* = 0$, $D^* = \Delta$; that is, the risk is not insured, or (ii) $A^* > 0$, $D^* = D(A^*) > 0$, for one or several levels of internal capital ($A^* > 0$), or both (i) and (ii). The following proposition shows, somewhat surprisingly, that the optimal contract is actually unique:

Proposition 4: *There is a unique optimal contract; that is, there are unique levels of internal capital, $A^* \geq 0$, and deductible, $D^* > 0$, such that the stop-loss contract with deductible defined in Proposition 1 is the optimal contract.*

As shown in the proof of the proposition, the uniqueness of the optimal contract follows from the fact that the Hessian of F, as a function of D and A, is strictly negative definite at any point where the FOCs for optimality are satisfied. This property of the Hessian is in general not sufficient to imply uniqueness for multivariate functions, but because the optimal D is unique as a function of A and depends smoothly on A (see Proposition 1) it turns out to be sufficient in this case.

We next analyze when the optimal contract is interior. We use the following terminology: If there is a solution with $A^* > 0$, then insurance is said to be optimal, whereas if the only solution is the one with $A^* = 0$, then insurance is said to be suboptimal. The following proposition characterizes when insurance is optimal:

Proposition 5: *Define*

$$\xi = \frac{1}{2}\frac{E|u'(-L) - \mu_{u'}|}{\mu_{u'}}. \tag{9}$$

Then,

1. *if $\delta > \xi$, insurance is suboptimal, whereas*
2. *if $\delta < \xi$, insurance is optimal.*

Thus, in addition to the cost of internal capital, δ, both risk distributions and insuree preferences (through (9)) are in general crucial in determining whether insurance is optimal for a specific insuree.

If, for a specific risk, insurance is suboptimal for *all* risk-averse agents, the risk is said to be *uninsurable*. It turns out that insurability only depends on the cost of internal capital, not on the specific risk distribution, as shown by the following proposition:

Proposition 6: *If $\delta \geq 1$ every risk is uninsurable, whereas if $\delta < 1$ no risk is uninsurable.*

The case $\delta \geq 1$ is obviously extreme since it implies that the insuree's premium is higher than the maximum insurance claim. It is unsurprising that in this case, risks are always uninsurable. But Proposition 6 also provides a positive case for insurance in that it shows that as long as there is any chance that the claim is higher than the premium, insurance will be optimal for some very risk-averse insurees. This is the interesting part of the proposition.

If insurance is optimal, the optimal contract implies a positive probability for insurer default. This is, of course, trivial since $A^* < \infty$ and, per assumption, the loss distribution has support on the whole of the positive real axis. However, it is less trivial that this property generalizes to situations where the loss distribution has bounded support, as shown by the following proposition:

Proposition 7: *If $\Phi(\Gamma) = 0$ for some $\Gamma < \infty$, that is, there is a finite upper bound on losses, then $D^* + A^* < \Gamma$; that is, the optimal insurance contract leads to a strictly positive risk of insurer default.*

The result is dual to the classical result of positive deductibles (Arrow 1963): the optimal contract cuts off payments for small losses through the deductible, as well as for large losses through insurer default. This new result is at first sight less intuitive than the result on deductibles though. The positive deductible is easily motivated, since the utility cost for the insuree of a very small loss is a second-order effect compared to the cost of lowering the deductible, which is of first order. Therefore, having a zero deductible is never optimal. In other words, the benefits of decreasing the deductible toward zero occur in the part of the state space where they matter the least to the insuree, but the costs are incurred over the whole state space, through a higher premium.

Such an argument is clearly not valid for the level of internal capital. In fact, increasing A, such that $A + D$ approaches Γ decreases the risk in exactly the states that matter the most, close to Γ. The way to see that it can never be optimal to choose a contract such that $A + D = \Gamma$, as shown in the proof of the proposition, is to focus on the deductible, not on the capital. Indeed, as shown in the proposition, the deductible is "too high," when chosen such that $A + D = \Gamma$. By decreasing the deductible, more insurance is provided in the high marginal utility states in which losses are higher than the deductible, whereas the cost of an increased premium decreases the utility in the low marginal utility states, below the deductible. When there is a positive probability for default, there is also a third effect, namely, that the higher premium decreases the utility in the worst states of default (when $L > A + D$), but when the starting point is one in which default does not occur, $A + D = \Gamma$, this effect is of second order, and a slight decrease of D is therefore always optimal. Thus, the result.

A standard argument for why insurer default may occur is that under diversification of a finite number of risks, there is still some residual risk that there will be so many bad outcomes that there is not enough internal capital to cover all claims. Thus, an insurer may default as a result of less than complete diversification. The argument above does not depend on diversification, since there is only one risk insured, and thus provides another rationale for why insurer default in some states of the world may be (*ex ante*) optimal.

The comparative statics of A^* and D^* with respect to changes in cost of internal capital, δ, are in general ambiguous: $dA^*/d\delta \lesseqgtr 0$, and $dD^* d\delta \lesseqgtr 0$. An example is given in the next section. However, in the case when the optimal deductible is small, an increase in δ can be shown to always *decrease* the optimal level of internal capital, in line with the intuition that higher costs of internal capital make insurance less attractive. Similar results are obtained when the cost of internal capital, δ, is high. Both effects follow from the following result

Proposition 8: *If $\Phi(D^*) + \delta \geq 1$, then the optimal amount of internal capital decreases in its cost, that is, $\frac{dA^*}{d\delta} < 0$.*

DISCRETE DISTRIBUTIONS

In the previous analysis, absolute continuity of the distribution functions was assumed. When loss distributions are discrete, somewhat different results may arise. We first study the simplest case, in which the risk that has a two-point distribution, $L \in Be(\Gamma, p)$, that is, when L takes on value $\Gamma > 0$ with probability p and 0 with probability $1 - p$. This case is especially simple, since the optimization problem in Proposition 1 is concave in this case, allowing for a complete characterization of the solution. Of course, it is without loss of generality that we assume that one of the outcomes is 0, since any positive loss that occurs for sure will be internalized by the insuree, who would never choose to "insure" against a certain loss when internal capital is costly. We note that in this case, it is sufficient for the utility function, u, to be defined on the interval $[-\Gamma,0]$, for the problem to be well defined.

Since the payout is never higher than $\Gamma - D$ regardless of A, and capital is costly, it will never be optimal to reserve more capital than $\Gamma - D$, that is, $A + D \leq \Gamma$. Furthermore, since the payout is the same in all states of the world for all deductibles $D \leq \Gamma - A$ (in the bad state the insurer defaults and pays A; in the good state no payments are made), we will without loss of generality assume that $D = 0$. The technical reason why the optimal D may not be unique, nor strictly positive, in this case—in contrast to what is implied by Proposition 1—is that probability density function of L is not strictly positive close to 0, as assumed previously, and this makes the FOC degenerate.

We recall that the insuree's absolute risk aversion at wealth level x is defined by $ARA(x) = -u''(x)/u'(x)$. We then have

Proposition 9: *Let $L \sim Be(\Gamma,p)$. Then the optimal amount of internal capital, A^*, is unique and satisfies*

1. $0 \leq A^* < \Gamma$.
2. *Insurance is suboptimal, $A^* = 0$, if and only if* $\frac{u'(-\Gamma)}{u'(0)} \geq \frac{1-p}{p} \times \frac{\delta+p}{1-\delta-p}$.
3. A^* *is increasing in* Γ, $\frac{dA^*}{d\Gamma} > 0$.
4. *For an insuree with decreasing absolute risk aversion (DARA), the value of the insurer's option to default is increasing in* Γ, $\frac{dA^*}{d\mu_Q} > 0$.
5. *For an insuree with increasing absolute risk aversion (IARA), the value of the insurer's option to default is decreasing in* Γ, $\frac{dA^*}{d\mu_Q} < 0$.
6. *Given that,* $A^* > 0$, *then* $\frac{dA^*}{d\delta} > 0$ *if and only if*

$$ARA\left((1-p-\delta)A^* - \Gamma\right) - ARA\left((-p-\delta)A^*\right) > \frac{1}{A^*(p+\delta)(1-p-\delta)}.$$

Proposition 9 thus shows that for this specific type of discrete distribution, the optimal amount of internal capital is still unique, that more internal capital will be held for larger risks, and that depending on the risk aversion of the insuree, the value of the insurer's option to default can be either larger or smaller for larger risks. Furthermore, the proposition shows that full insurance is never optimal, since condition 1 implies that $A^* < \Gamma$. This is in line with the result in the previous section of a positive probability for insurer default. The interpretation is slightly different here though, because of the degeneracy of the chosen deductible. Specifically, our convention is that $D = 0$, and $A^* < \Gamma$ then implies a positive probability of default. However, an equivalent contract (leading to the same compensation in all states of the world) is to choose $D = \Gamma - A^*$, which never leads to default.

Condition 3 shows that increasing the severity of the risk, in the sense of the magnitude of the bad outcome leads to a higher level of internal capital. This result depends on the specific definition of "increased risk." Another definition of risk is in the sense of second-order stochastic dominance. We study the risks $L \in Be(\Gamma/p, p)$, $\Gamma > 0$, $0 < p < 1$, which when p varies all have expected value $E[L] = \Gamma$, but have higher risk for higher p in terms of second-order stochastic dominance. It is easy to check that $dA^*/dp \lesseqgtr 0$, depending on the utility and risk distribution. Thus, the relationship between risk and level of internal capital is ambiguous when second-order stochastic dominance is used even in this simple case.

Condition 6 implies that nonincreasing absolute risk aversion is sufficient for the amount of internal capital to be decreasing in the cost of holding internal capital (since $A^* < \Gamma$, so $(1 - p - \delta)A^* - \Gamma < (-p - \delta)A^*$). It is straightforward to construct examples with DARA preferences such that $dA^*/d\delta < 0$.[5]

For general discrete N-point risk distributions, the results become more complex. The multiplicity of optimal contracts also hold with N-point distributions, as long as the support of the distribution does not lie within D^* and $D^* + A^*$. Specifically, if the support of L lies on $\ell_1 < \ell_2 < \ldots < \ell_N{}'$ then if $\ell_i < D^* < D^* + A^* < \ell_{i+1}$ for some i, then any other $D' \neq D$ such that $\ell_i < D' < D' + A^* < \ell_{i+1}$ leads to identical payouts in all states of the world and is therefore also optimal. If $\ell_i < D^* < \ell_{i+1} < D^* + A^*$, on the other hand, any change in D^* also changes the claim function in some states of the world.

MULTIPLE POLICYHOLDERS

As mentioned, we expect similar results to hold more generally when there are multiple policyholders, who insure risks that are not perfectly correlated, at least in the situation where the policyholders are "similar."

We study the case where there are N policyholders with identical utility functions, each facing a risk, L_i. The joint probability density function for these N risks is $\phi(x_1, x_2, \ldots, x_{N-1}, x_N)$, and we assume that the density function is strictly positive on $[0, \Gamma]^N$, where $0 < \Gamma \leq \infty$, covering both the case of bounded and unbounded risk distributions. The risk distributions are symmetric in that

$$\varphi(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_N) = \varphi(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_N), \quad 1 \leq i, j \leq N.$$

There is one insurer, who chooses internal capital, A, and N insurance contracts to offer to the insurees, $I_1, \ldots, I_N$, and policy premia, $P_1, \ldots, P_N$, such that

$$0 \leq I_i(L_1, \ldots, L_N) \leq L_i, \quad 1 \leq i \leq N, \tag{10}$$

$$A \geq \sum_{i=1}^{N} I_i(L_1, \ldots, L_N), \tag{11}$$

$$0 \leq P_i, \quad 1 \leq i \leq N, \tag{12}$$

[5] An example is given by $\delta = 0.6$, $p = 0.02$, $\Gamma = 1$, $u(x) = -(x + 1.1)^{-5}$, for which $A^* = 0.6335$ and $\frac{dA^*}{d\delta} = -0.0058$.

$$\sum_{i=1}^{N} P_i = \sum_{i=1}^{N} E[I_i(L_1, \ldots, L_N)] + \delta A. \tag{13}$$

These conditions generalize (1,2) in the single policyholder case. Condition (10) ensures that each policyholder receives a nonnegative payment, and never more than actual losses. Note here that the contract is allowed to depend not only on the losses of the actual policyholder, but also on the losses of all other policyholders. This is obviously a crucial property when there may not be sufficient capital for the insurer to cover all losses. Condition (11) ensures that the insurer never pays more than existing capital, (12) that insurance premia are always nonnegative, and (13) that pricing is competitive, that is, that the risk-neutral insurer earns no surplus.

The insurer designs a set of contracts and premia to optimize the total (symmetrically weighted) utility of the insurees; that is, it solves

$$\max_{A, I_1, \ldots, I_N, P_1, \ldots, P_N} \quad \sum_{i=1}^{N} E\left[u(-L_i + I_i - P_i)\right], \quad \text{s.t.}(10-13). \tag{14}$$

We then have

Proposition 10: *The optimal contracts offered by the insurer, $I_i\ (L_1, \ldots, L_N)$, $1 \le i \le N$, are on the form*

$$I_i(L_1, \ldots, L_N) = \max(L_i - D^*, 0), \quad when \sum_i \max(L_i - D, 0) \le A^*,$$

$$I_i(L_1, \ldots, L_N) = \max(L_i - D^* - Q, 0), \quad when \sum_i \max(L_i - D, 0) > A^*.$$

where the deductible $D^* > 0$ and internal capital $A^* \ge 0$. Here, $Q > 0$ is chosen such that $\sum_i \max(L_i - D^* - Q, 0) = A^*$. The premium is the same for all policyholders, $\mathrm{P}_i = \mathrm{P}^*$. The optimal level of internal capital, A^*, and corresponding deductible, D^*, are such that

$$N(\Gamma - D^*) > A^*. \tag{15}$$

Again, the optimal contract contains a deductible, above which policyholders are fully insured in case total claims do not exceed internal capital ($\sum_i \max(L_i - D, 0) \le A^*$).

If total claims exceed internal capital ($\sum_i \max(L_i - D, 0) \le A^*$), then the insurer defaults, and distributes its capital among claimants such that the highest losses receive the highest priority in claim payments. This rule arises naturally with risk-averse investors because of increasing marginal utilities of net losses. Finally, (15) implies a strictly positive probability for insurer default, just as in the case with a single policyholder.

The symmetric setting (over preferences, risk distributions, and utility weights) is crucial for the strong results in this section. In the symmetric setting, the sole purpose of the insurance contracts is to allow a risk transfer from the risk-averse policyholders to the risk-neutral insurer. In an asymmetric setting, risk transfers between policyholders also arise, which makes the problem much harder. In a homogeneous market, where policyholders and risks are similar, our model may therefore provide a decent approximation, whereas in heterogeneous markets a richer model is needed. An example of the latter situation may be a multiline insurer, for which risk distributions may vary significantly across lines. We leave such considerations for future research.

CONCLUDING REMARKS

The introduction of costly internal capital leads to a rich set of implications in an otherwise standard insurance model, linking the insurance literature to the corporate finance literature in which the assumption about costly capital is standard. A natural future extension of our model would be to introduce multiple insurance lines, insured by the same insurance company. In such a setting, risk distributions would be asymmetric. What is the optimal contract in such a setting, and how does this contract compare with the contracts seen in practice? How do diversification benefits affect the optimal level of internal capital in this case? These are important open questions, which we hope to address in future work.

APPENDIX

Proof of Proposition 1: Form of optimal contract: We first show that any optimal contract must be on the form given in the proposition. The general set of contracts I can be defined as the set of measurable compensation functions satisfying the constraints $0 \le I(x) \le \min(x, A)$. We define the subset of contracts $I_0 \subset I$ that are piecewise continuous with a finite number of discontinuities and note that this set is dense in I under the L^1-metric (with distance function $d(a, b) = \int |a(x) - b(x)| dx$). Furthermore, the operation $I \mapsto Eu[-x + I(x) - P]$ is (L^1) continuous, so if there is an optimal contract in I_0, this is also a globally optimal contract (within L^1).

The variational problem within I_0 is

$$U = \max_{I(x) \in I_0} \int_0^\infty u\left(-x + I(x) - \int_0^\infty I(y)\varphi(y)dy - \delta A\right)\varphi(x)dx,$$

so that $0 \le I(x) \le \min(x, A)$. The first variation is

$$\left.\frac{\delta U}{\delta I}\right|_x = u'\left(-x + I(x) - \int_0^\infty I(y)\varphi(y)dy - \delta A\right) - \Gamma,$$

where $\Gamma = \int_0^\infty u'(-x + I(x) - P)dx$. The condition implies that there are three possibilities: (1) that no constraints bind, in which case the first variation is zero

$u'(-x + I(x) - P) = \Gamma$; (2) that the nonnegativity constraint binds, $I(x) = 0$, in which case the first variation is negative $u'(-x + I(x) - P) < \Gamma$; and (3) that the upper constraint binds, $I(x) = \min(x, A)$, in which case the first variation is positive $u'(-x + I(x) - P) > \Gamma$. We define the three sets B_1, B_1, and B_3, in which the first variation is zero, negative, and positive, respectively.

Since u' is strictly decreasing, the total position of the insuree, $-x + I(x) - P$, under the first condition must be constant, that is, $I(x) = x - D$, for some $D \geq 0$. Furthermore, if this position is feasible for x_1 and $x_2 > x > x_1$, then it is also feasible for all $x \in [x_1, x_2]$, so the region in which the contract takes this form, B_1, must be an interval. Indeed, it must be the interval $[D, D + A]$, since the lower constraint becomes binding at $x = D$ and the upper constraint at $x = D + A$.

For $x < D$, we are in the second situation above, where the nonnegativity constraint must be binding, so that $I(x) = 0$, and, finally, for $x > D + A$, we are in the third situation, and the upper constraint must be binding so that $I(x) = A$ (since $x > D + A$). Thus, $B_2 = [0, D$, and $B_3 = (D + A, \infty)$. So, any contract in I_0 can be improved upon by choosing the stop-loss contract with some deductible, D. We call the set of contracts on this form $I_1 \subset I_0$. Therefore, if there is a uniquely optimal contract, I, in I_1, then this is indeed the globally optimal contract in I.

Existence and uniqueness: Now, let us focus on contracts $I \in I_1$ and define $F = Eu(-L + I(L) - P)$, which is obviously smooth as a function of D. It follows that the partial derivative, $\frac{\partial F}{\partial D}$, is given by

$$\frac{\partial F}{\partial D} = Z\alpha - Z(1 - Z)u'(-D - P) + Z\beta, \tag{A1}$$

where

$$P = A\Phi(D + A) - DZ + \int_D^{D+A} x\varphi(x)dx + \delta A,$$

and

$$Z = \Phi(D) - \Phi(D + A),$$

$$\alpha = \int_{D+A}^{\infty} u'(-x + A - P)\varphi(x)dx,$$

$$\beta = \int_0^D u'(-x - P)\varphi(x)dx.$$

At $D = 0$, we have

$$\left.\frac{\partial F}{\partial D}\right|_{D=0} = (1 - \Phi(A))(\alpha - \Phi(A)u'(-P)) = (1 - \Phi(A))\left(\int_A^{\infty} u'(-x - P) - u'(-P)dx\right) > 0.$$

Furthermore, for large D we have

$$\frac{\partial F}{\partial D} = Z(-u'(-D-P) + \int_0^\infty u'(-x-P)\varphi(x)dx + o(1)) < 0.$$

Here, $o(1)$ is a higher order term, meaning that $\lim_{D\to\infty} o(1) = 0$. Thus, there must be at least one internal optimum, which will be characterized by the FOC.

What remains is to prove uniqueness. Consider a D, such that the FOC is satisfied. The second-order condition at such a point is

$$\begin{aligned}\frac{\partial^2 F}{\partial D^2} = {} & \frac{\partial Z/\partial D}{Z} \times \frac{\partial F}{\partial D} \\ & + Z\left((1-Z)^2 u''(-D-P) - \frac{\partial P}{\partial D}\left(\int_{D+A}^\infty u''(-x+A-P)\varphi(x)dx + \int_0^D u''(-x-P)\varphi(x)dx\right)\right) \\ & = 0 + Z\left((1-Z)^2 u''(-D-P) + Z\left(\int_{D+A}^\infty u''(-x+A-P)\varphi(x)dx + \int_0^D u''(-x-P)\varphi(x)dx\right)\right) \\ & < 0,\end{aligned}$$

where we have used that $\frac{\partial P}{\partial D} = -Z$, and $Z > 0$. Thus, any point where the FOC is satisfied is a maximum, and consequently there is exactly one such point.

Continuity of D as a function of A follows from the smoothness of F as a function of D and A, that $\partial^2 F/\partial D^2 < 0$ at the optimum, and the implicit function theorem. In fact, F is smooth enough to guarantee that $D(A)$ is continuously differentiable, which we will use in the proof of Proposition 4. We are done.

Proof of Proposition 2: For A close to zero, $Z = A\varphi(D)(1 + O(A))$, and $P = O(A)$. Here, $O(A)$ represents a function such that $|O(A)| \le CA$ for some constant $C > 0$.

The FOC (A1) can now be written

$$\begin{aligned}\frac{\partial F}{\partial D} & = Z\alpha - Z(1-Z)u'(-D-P) + Z\beta \\ & = A\varphi(D)(1+O(A))\left(\int_D^\infty u'(-x)\varphi(x)dx - u'(-D) + \int_0^D u'(-x)\varphi(x)dx + O(A)\right) \\ & = 0,\end{aligned}$$

leading to

$$u'(-D) = \mu_{u'} + O(A).$$

The continuity and invertibility of u', and the strict negativity of u'', now implies the result.

Proof of Proposition 3: From (2) and Proposition 1, it follows that

$$P = \int_{D}^{D+\mathrm{A}} x - D)\varphi(x)dx + \int_{D+\mathrm{A}}^{\infty} \mathrm{A}\varphi(x)dx + \delta \mathrm{A} \tag{A2}$$

$$= \int_{D}^{D+A} x\varphi(x)dx - D(\Phi(D) - \Phi(D+A)) + A\Phi(A+D) + \delta A, \tag{A3}$$

and 1–3 follow immediately from this representation. Furthermore, 4–7 follow from taking partial derivatives of first and second order of this expression.

Proof of Proposition 4: From Proposition 1, we know that any optimal contract must be a stop-loss contract with a deductible. Thus, if we can show that there is a unique $(A, D) \in R^2_+$, among all such contracts that is optimal, we are done.

From the proposition, we know that for each possible A, $D(A)$ is unique. It also follows immediately from the costly internal capital, δA, $\delta > 0$, that any optimal solution must have $A^* \leq \bar{A}$, where $u(-\delta\bar{\mathrm{A}}) = Eu(-L)$. Thus, the maximal contract must lie in the compact set $\{(A, D^*(A)) : 0 = A \leq \bar{A}\}$.

Since the objective function, $F(A,D)$ belongs to $C^2(R^2_+)$, an interior optimal solution must necessarily satisfy the FOC (5,6), but since the problem is not concave, it is *a priori* unclear that an interior optimum is unique. However, if it can be shown that at any point satisfying the FOC, the problem is locally concave in the sense that the Hessian,

$$H = \begin{bmatrix} \dfrac{\partial^2 F}{\partial A^2} & \dfrac{\partial^2 F}{\partial A \partial D} \\ \dfrac{\partial^2 F}{\partial A \partial D} & \dfrac{\partial^2 F}{\partial D^2} \end{bmatrix},$$

is strictly negative definite, then any interior optimum is indeed unique. Otherwise, the smooth function $G(A) = F(A, D^*(A))$ would contain multiple maxima, without any local minimum, since $0 = G\prime(A) = \frac{\partial F}{\partial A} + \frac{\partial F}{\partial D}\frac{dD^*}{dA}$ can only occur at points where both FOCs for F are satisfied ($\frac{\partial F}{\partial D} = 0$, as is then also $\frac{\partial F}{\partial A}$) and they are always local maxima. Such a situation is of course impossible, and there can therefore only be—at most—one interior optimum.

Furthermore, if there is an interior maximum, $(A^*, D(A^*))$, with $A^* > 0$, then G is necessarily increasing for $A < A^*$, so $(0,D(0))$ can not be a maximum. Similarly, if $(0,\Delta)$ is a maximum, then $G(A)$ is nonincreasing in a neighborhood of $A = 0$. This can either occur if $G'(0) < 0$, or if $G'(0) = 0$, but in either case a similar argument as above implies that the no-insurance outcome is the unique maximum in this case. Thus, it is sufficient to prove that the Hessian is strictly negative definite at any point where the FOCs are satisfied.

A second-order Taylor expansion of F at a point where $\partial F/\partial A = 0$ yields:

$$\frac{\partial^{2F}}{\partial A^2} = Q^2 c + (Q-1)^2 \hat{a} + ZQ^2 \hat{m},$$

$$\frac{\partial^2 F}{\partial D^2}, = Z^2 c + Z^2 \hat{a} + (Z-1)^2 Z\hat{m},$$

$$\frac{\partial^2 F}{\partial A \partial D} = ZQc + Z(Q-1)\hat{a} + (Z-1)ZQ\hat{m},$$

where

$$\hat{a} = \int_{D+A}^{\infty} u''(-x + A - P)\varphi(x)dx < 0,$$

$$\hat{m} = u''(-D - P) < 0,$$

$$c = \int_{0}^{D} u''(-x - P)\varphi(x)dx < 0,$$

$$Q = \delta + O(D + A) > 0,$$

and, as before, $Z = \Phi(D) - \Phi(A+D)$.

By inspection, $\frac{\partial^2 F}{\partial A^2} < 0$ and $\frac{\partial^2 F}{\partial D^2} < 0$ (which we already proved in Proposition 2). Furthermore,

$$det[H] = \frac{\partial^2 F}{\partial A^2} \times \frac{\partial^2 F}{\partial D^2} - \left(\frac{\partial^2 F}{\partial A \partial D}\right)^2 = Z(\hat{m}\hat{c}Q^2 + \hat{a}\hat{c}Z + \hat{a}\hat{m}(Q + Z - 1)^2) > 0,$$

and together this implies that H is indeed strictly negative definite. Thus, the result follows.

Proof of Proposition 5: Sufficiency: We calculate $\frac{\partial F}{\partial A}$ at $D = \Delta$, $A = 0$,

$$\begin{aligned} \left.\frac{\partial F}{\partial A}\right|_{A=0,D=\Delta} &= \alpha - (\delta + \Phi(D + A))(\alpha + Zu'(-D - P) + \beta), \\ &= \int_{\Delta}^{\infty} u'(-x)\varphi(x)dx - (\delta + \Phi(\Delta))u'(-\Delta) \\ &= \int_{\Delta}^{\infty} (u'(-x) - \mu_{u'})\varphi(x)dx - \delta u'(-\Delta) \\ &= \int_{\Delta}^{\infty} |u'(-x) - \mu_{u'}|\varphi(x)dx - \delta u'(-\Delta) \\ &= \int_{\Delta}^{\infty} |u'(-x) - \mu_{u'}|\varphi(x)dx - \delta u_{u'}. \end{aligned} \tag{A4}$$

Now, from (4), we have

$$\begin{aligned} 0 &= \int_{0}^{\infty} (u'(-x) - \mu_{u'})\varphi(x)dx \\ &= \int_{\Delta}^{\infty} (u'(-x) - \mu_{u'})\varphi(x)dx + \int_{0}^{\Delta} (u'(-x) - \mu_{u'})\varphi(x)dx, \\ &= \int_{\Delta}^{\infty} (u'(-x) - \mu_{u'})\varphi(x)dx - \int_{0}^{\Delta} (u'(-x) - \mu_{u'})\varphi(x)dx, \end{aligned}$$

implying that

$$\begin{aligned} E[|u'(-L) - \mu_{u'}|] &= \int_{\Delta}^{\infty} |u'(-x) - \mu_{u'}|\varphi(x)dx + \int_{0}^{\Delta} |u'(-x) - \mu_{u'}|\varphi(x)dx \\ &= 2\int_{\Delta}^{\infty} |u'(-x) - \mu_{u'}|\varphi(x)dx. \end{aligned}$$

Plugging this into (A4) yields

$$\left.\frac{\partial F}{\partial A}\right|_{A=0,D=\Delta} = \frac{1}{2}E[|u'(-L) - \mu_{u'}|] - \delta\mu_{u'}, \tag{A5}$$

and if this is positive, a small amount of insurance is indeed better than no insurance, so insurance is optimal.

Necessity: Now, assume that $\frac{1}{2}E[|u'(-L)\mu_{u'}|] < \delta\mu_{u'}$. We use a similar argument as in the proof of Proposition 4. First, note that $\frac{\partial F}{\partial D}\Big|_{A=0} = 0$. A Taylor expansion of $G(A)$ in a neighborhood of $A = 0$ yields

$$G(\Delta A) = \frac{\partial F}{\partial A}\Delta A + \frac{\partial F}{\partial D}D^{*'}(0)\Delta A + o(\Delta A) = \left(\frac{1}{2}E[|u'(-L) - \mu_{u'}|]\delta\mu_{u'}\right)\Delta A + o(\Delta A) < 0.$$

A similar argument as in the proof of Proposition 4 implies that G is decreasing for all A, and thus that no insurance is the optimal solution. We are done.

Proof of Proposition 6: $\delta \geq 1$: It is trivial that insurance is always suboptimal in this case: Any insurance contract with $A > 0$ will lead to strictly first-order stochastically dominated payoffs compared with no insurance, since the price for insurance is at least as high as what can ever be claimed, and in some states higher.

$\delta < 1$: The result follows almost immediately from the following lemma.

Lemma: *sup* : $\frac{E[|\bar{Z}-\mu_Z|]}{\mu_Z} = 2$, *where the supremum is taken over all nonnegative random variables with infinitely differentiable probability density functions.*

Proof: From the triangle inequality it follows that $E[|Z - \mu_Z|] \leq E[|Z|] + \mu_Z = 2\mu_Z$, so 2 is an upper bound. Now, let Z have a two-point distribution with probability 1 $- \varepsilon$ that $Z = 0$ and probability ε that $Z = \frac{\bar{Z}}{\varepsilon}$, $\bar{Z} > 0$. Then, $\mu_Z = \bar{Z}$, and $E[|Z - \mu_Z|] = (1 - \varepsilon)\mu_Z + \varepsilon(\frac{\bar{Z}}{\varepsilon} - \mu_\varepsilon) = 2(1 - \varepsilon)\mu_Z$. Thus, by choosing ε arbitrarily close to zero, one can get arbitrarily close to the upper bound. The only technicality is that the distribution is not smooth. However, since the infinitely differentiable distribution functions form a dense subset of the set of distribution functions, and the operations involved are continuous, the supremum is 2 also for this subset.

Now, given the risk L with density function ϕ, it is clear that a smooth increasing function u' can be chosen so that the distribution of $Z \stackrel{\text{def}}{=} u'(-L)$ is arbitrarily close to the sequence of two-point distributions in the lemma, that is, such that $E[|u'(-L) - \mu_{u'}|] \geq 2(1 - \varepsilon)\mu_{u'}$, for arbitrary $\varepsilon > 0$. But when plugging this into (A5), it then follows that

$$\frac{\partial F}{\partial A}\Big|_{A=0,\,D=\Delta} = \frac{1}{2}E[|u'(-L) - \mu_{u'}|] - \delta\mu_{u'} \leq \left(\frac{1}{2}2(1 - \varepsilon) - \delta\right)\mu_{u'} > 0,$$

for small enough ε, so the risk is indeed insurable. *Q.E.D*

Proof of Proposition 7: It is clear that it can never be optimal to choose the deductible and internal capital such that $D + A > \Gamma$, since by choosing $A = \Gamma - D$ the contract makes the same payments in all states but the premium is lower. Thus, it is sufficient to focus on contracts such that $D + A = \Gamma$.

Now, assume that the optimal deductible and internal capital is such that $D^* + A^* = \Gamma$. The FOC with respect to the deductible is in this case:

$$\begin{aligned}\frac{\partial F}{\partial D} &= \Phi(D)\left((1-\Phi(D))u'(-D-P) - \int_0^D u'(-x-P)\varphi(x)dx\right)\\ &= -\Phi(D)\int_0^D (u'(-D-P) - u'(x-P))\varphi(x)dx < 0.\end{aligned}$$

Thus, a better contract can be constructed by decreasing the deductible, leading to a contradiction, and therefore no optimal contract can have $D + A \geq \Gamma$.

Proof of Proposition 8: Taking the differential of

$$\begin{pmatrix} \frac{\partial F}{\partial A} \\ \frac{\partial F}{\partial D} \end{pmatrix}$$

around A^*, D^*, δ yields

$$\begin{pmatrix} \frac{\partial F}{\partial A} \\ \frac{\partial F}{\partial D} \end{pmatrix}_{A^*+dA,\, D^*+dD;\delta+d\delta} - \begin{pmatrix} \frac{\partial F}{\partial A} \\ \frac{\partial F}{\partial D} \end{pmatrix}_{A^*,\, D^*;\delta} = H\begin{pmatrix} dA \\ dD \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 F}{\partial A \partial \delta} \\ \frac{\partial^2 F}{\partial D \partial \delta} \end{pmatrix} d\delta,$$

where H is the Hessian matrix defined in Proposition 4.

To keep the FOC, the differential must be zero, so

$$\begin{pmatrix} dA \\ dD \end{pmatrix} = -H^{-1}\begin{pmatrix} \frac{\partial^2 F}{\partial A \partial \delta} \\ \frac{\partial^2 F}{\partial D \partial \delta} \end{pmatrix} d\delta.$$

Using the fact that H is negative definite at the optimum, implying that $det[H] > 0$, calculating

$$\frac{\partial^2 f}{\partial A \partial \delta} = -\mathrm{u}'(-\mathrm{D}-\mathrm{P}) + A(Q\hat{c} + (Q-1)\hat{a} + Q\hat{m}),$$

$$\frac{\partial^2 F}{\partial D \partial \delta} + AZ(c + \hat{a} + (Z-1)\hat{m}),$$

and plugging in the elements of H, we get

$$\frac{dA}{d\delta} = -\frac{1}{det[H]}\left(\frac{\partial^2 F}{\partial D^2} \times \frac{\partial^2 F}{\partial A \partial \delta} - \frac{\partial^2 F}{\partial D \partial A} \times \frac{\partial^2 F}{\partial D \partial \delta}\right)$$

$$= -\frac{Z}{det[H]}\ \hat{m}^2 Q(1-Z)^3 + \hat{m}\hat{c}Q(1+Z-Z^2) + \hat{a}\hat{m}((Z+Q-1) + QZ(1-Z))$$
$$+ u'(D+P)\left(-\frac{\partial^2 F}{\partial D^2}\right).$$

Here, all terms are defined as in the proof of Proposition 4.

We note that all terms within the parentheses are immediately strictly positive, except for $Z + Q - 1 = \Phi(D) - \Phi(D+A) + \delta + \Phi(D+A) - 1 = \Phi(D) + \delta - 1$. However, under the condition of the proposition, this term is also weakly positive, so the expression within the parentheses is positive, and $\frac{dA}{d\delta}$ is therefore strictly positive.

Proof of Proposition 9: 1 and 2. Clearly, since $\delta > 0$, choosing $A = \Gamma$ always dominates choosing $A > \Gamma$, as the insurance payoffs are identical in both states of the world, but the cost of internal capital is higher if $A > \Gamma$ than if $A = \Gamma$. Thus, the solution, which is unique, since the objective function is strictly concave, must lie in $[0,\Gamma]$. Furthermore, it is clear that if $\delta + p \geq 1$, then $A^* = 0$ since self-insurance is optimal in this case. The results are trivial in this case and we therefore proceed with the case when $\delta + p < 1$.

Define $q = \delta + p$. The FOC from (5) is

$$(1-p)qu'(-qA^*) = p(1-q)u'((1-q)A^* - \Gamma),$$

which, when the function $b_\Gamma(A) \stackrel{\text{def}}{=} \frac{u'(-qA)}{u'((1-q)A-\Gamma)}$ is defined, is equivalent to

$$b_\Gamma(A^*) = \frac{p}{1-p} \times \frac{1-q}{q}, \tag{A6}$$

where, since $q > p$, the right-hand side is strictly less than 1. Now, since u is strictly concave and twice continuously differentiable, it follows that $b_\Gamma(A)$ is strictly increasing in A. Also, $b_\Gamma(0) = \frac{u'(0)}{u'(-\Gamma)}$, and if $b_\Gamma(0) \leq \frac{p}{1-p} \times \frac{1-q}{q}$, it must then be that $A^* = 0$, which immediately implies 2. Also, since $b_\Gamma(\Gamma) = 1 > \frac{p}{1-p} \times \frac{1-q}{q}$, the maximum must indeed be realized for $A^* < \Gamma$, so 1 follows. We also note that the monotonicity of the utility function, u, obviously implies that b_Γ is positive.

3. Define $V = b_\Gamma - \frac{p}{1-p}\frac{1-q}{q}$. Then, the FOC is $V(A^*, \Gamma) = 0$. A Taylor expansion around (A^*, Γ) yields $dV = \frac{\partial V}{\partial A^*}dA^* + \frac{\partial V}{\partial \Gamma}d\Gamma$. From the previous argument, we know that $\frac{\partial V}{\partial A^*} = b'_\Gamma > 0$. So, it is sufficient to show that $\frac{\partial b_\Gamma(A)}{\partial \Gamma} < 0$ for the result to follow, since $\frac{dA^*}{dV} = -\frac{\partial V \partial \Gamma}{\partial V \partial A^*}$. Differentiating V with respect to Γ leads to $\frac{\partial V}{\partial \Gamma} = -ARA((1-q)A - \Gamma)b_\Gamma < 0$, and the result follows.

4 and 5. We have $P_Q(A) = p(\Gamma - A)$ for $A < \Gamma$. Therefore, $\frac{dP_Q}{d\Gamma} = p\left(1 - \frac{dA^*}{d\Gamma}\right)$. So, if $\frac{dA^*}{d\Gamma} < 1$, then $\frac{dP_Q}{d\Gamma} > 0$.

From the proof of 3, it follows that

$$\frac{dA^*}{d\Gamma} = -\frac{\dfrac{\partial b_\Gamma}{\partial \Gamma}}{\dfrac{\partial b_\Gamma}{\partial A}} = \frac{ARA((1-q)A - \Gamma)b_\Gamma}{(q\,ARA(-q\,A) + (1-q)ARA((1-q)A - \Gamma))b_\Gamma}$$
$$= \frac{ARA((1-q)A - \Gamma)}{qARA(-q\,A) + (1-q)ARA((1-q)A - \Gamma)}.$$

Now, since $(1-q)A - \Gamma < -qA$, and for an agent with DARA preferences ARA is positive and decreasing, $\frac{dA^*}{d\Gamma}$ is therefore less than 1, and 4. follows. A similar argument for IARA preferences leads to 5.

6. Given that $V = b_\Gamma - \frac{p}{1-p}\frac{1-q}{q}$. Then, the FOC is $V(A^*, \delta) = 0$. A Taylor expansion around (A^*, δ) yields $dV = \frac{\partial V}{\partial A^*}dA^* + \frac{\partial V}{\partial \delta}d\delta$. From the previous argument, we know that $\frac{\partial V}{\partial A^*} = b'_\Gamma > 0$, and we get

$$\frac{\partial V}{\partial \delta} = \frac{p}{1-p}\frac{1}{q^2} + A^* \times ARA((1-q)A^* - \Gamma) \times \frac{u'(-A^*q)}{u'((1-q)A^* - \Gamma)} - A^* \times ARA(-q\,A^*)$$
$$\times \frac{u'(-A^*q)}{u'((1-q)A^* - \Gamma)} = \frac{p}{1-p}\frac{1}{q^2} + A^*\frac{p}{p-p}\frac{1-q}{q}\big(ARA((1-q)A^* - \Gamma)$$
$$-ARA(-q\,A^*)\big) = \frac{p}{1-p}\frac{1}{q^2}\left(1 + A^*q(1-q)(ARA((1-q)A^* - \Gamma) - ARA(-q\,A^*))\right).$$

For comparative static purposes, we set $dV = 0$, to get $\frac{dA^*}{d\delta} = -\frac{\frac{\partial V}{\partial \delta}}{b'\Gamma}$ and the condition in 6 is then indeed necessary and sufficient for $\frac{dA^*}{d\delta} < 0$. We are done.

Proof of Proposition 10: We first show that the solution to (14) must be symmetric, in that contracts and premia must be the same across policyholders.[6] Consider a proposed asymmetric optimal solution that charges $P_1, \ldots, P_N$ to the N policyholders, with the set of contracts, $I_1, \ldots, I_N$. Moreover, consider the realized losses $L_1 = x_1, \ldots, L_N = x_N$, which we arrange in a vector $\mathbf{x} = (x_1, \ldots, x_N)$. Because of the symmetry of loss

[6] Strictly speaking, contracts need to be asymmetric on a set of positive measure.

distributions, the probability density for these losses, $p = \phi(\mathbf{x})$ is the same as for any permutation of losses, $p = \phi(\mathbf{y})$, where $\mathbf{y}_i = \mathbf{x}_{V(i)}$, and $V : \{1, \ldots, N\} \to \{1, \ldots, N\}$ is a permutation (bijection). There are $M = N!$ such permutations, which we enumerate by $V_1, \ldots, V_M$.

We study the permutations of losses $\mathbf{y}^m$ such that $\mathbf{y}_j^m = \mathbf{x}_{V_m(j)}, m = 1, \ldots, M$, and therefore define $X_{i,m} = I(\mathbf{y}^m)_{V_m^{-1}(i)} - x_i$. $X_{i,m}$ thus defines claims minus losses for the policyholder who loses $\mathbf{x}_i$ in a specific permutation, whereas $P_{V_m(i)}$ defines the premium paid by that policyholder. A symmetric contract would have $X_{im} = X_{i,m'}$ for all i, m, and m', as well as $P_i = P_{i'}$ for all i.

The contribution to expected utility in (14) by all M loss realizations, $\mathbf{y}^1, \ldots, \mathbf{y}^M$, is $p \sum_{i=1}^N \sum_{m=1}^M u(X_{im} - P_{V_m(i)})$. Now, define $P = \frac{1}{N} \sum_{i=1}^N P_i$, and $X_i = \frac{1}{M} \sum_{m=1}^M X_{im}$. Jensen's inequality then implies that

$$p \sum_{i=1}^{N} \sum_{m=1}^{M} u(X_{im} - P_{V_m(i)}) < p \sum_{i=1}^{N} Mu\left(\frac{1}{M} \sum_{m=1}^{M} (X_{im} - P_{V_m(i)})\right) = pM \sum_{i=1}^{M} u(X_i - P). \quad \text{(A7)}$$

Equation (A7) shows that any set of asymmetric contracts can be strictly improved upon by creating a symmetric contract for all policyholders, that averages the payouts and premium across these policyholders. It is easy to check that if (10–13) hold for the original contract, the conditions also hold for the symmetric contract, so this contract is feasible. Finally, since $\mathbf{x}$ was arbitrary, the arguments holds for all losses. Thus, the optimal contract must be symmetric.

Restricting our attention to symmetric contracts, we can restate the optimization problem as

$$\max_{A, I, P} E\left[u(-L_1 + I(L_1) - P)\right], \quad \text{s.t.,}$$

$$0 \quad \leq I(L_1) \leq L_1,$$

$$A \geq \sum_{i=1}^{N} I(L_i),$$

$$P = E[I(L_1)] + \delta \frac{A}{N}.$$

This is very similar to the case with one policyholder, and the same variational conditions arise,

$$\left.\frac{\delta U}{\delta I}\right|_x = u'(-x + I(x) - P) - \Gamma,$$

which is equal to zero unless constraints bind, in which case the variation is negative if the nonnegativity constraint binds, and positive if the finite internal capital constraint binds.

An identical argument as in the proof of Proposition 1 immediately implies that the form of the contract in case of sufficient internal capital to cover all losses ($\sum_i max(L_i - D, 0) \leq A$) is $I_i = max(L_i - D, 0)$, for some $D > 0$ and $A \geq 0$.

The only extra dimension compared with the single policyholder case is that when there is not sufficient capital to cover all losses, $\sum_i max(L_i - D, 0) > A$, an additional question of how to distribute the capital among the claimants arises. Since marginal utilities are increasing in net losses, it follows that in case of insufficient capital, the optimal allocation is such that the maximum marginal utility among claimants is minimized, which is exactly what the rule $\max(L_i - D - Q, 0)$, where $\sum_i \max(L_i - D - Q, 0) = A$, achieves. This shows the second part of the contract.

It remains to show that the probability for default is strictly positive, that is, that $N(\Gamma - D^*) > A^*$. Of course, this is trivial when $\Gamma = \infty$, so we focus on the case with bounded support of loss distributions. As in Proposition 7 we use $F = Eu(-L_1 + I(L_1) - P)$, and an identical argument as in that proposition yields

$$
\begin{aligned}
ll \left.\frac{\partial F}{\partial D}\right|_{\Gamma - D = A/N} &= \Phi(D)Eu'(-L_1 + I(L_1) - P) - \Phi(D)u'(-D - P) \\
&= \Phi(D)\left(\int_0^D u'(-x - P)\varphi(x)dx + \Phi(D)u'(-D - P) - u'(-D - P)\right) \\
&= \Phi(D)\int_0^D (u'(-x - P) - u'(-D - P))\varphi(x)dx < 0,
\end{aligned}
$$

where $\Phi(D)$ is the complementary (marginal) distribution function, and $\phi(x)$ is the (marginal) probability density function of L_1. Thus, as in Proposition 7, this argument implies that the deductible is too high when $N(\Gamma - D^*) = A^*$, and the result therefore follows.

REFERENCES

Arrow, K., 1963, Uncertainty and the Welfare Economics of Medical Care, *American Economic Review*, 53: 1-42.

Arrow, K., 1971, *Essays in the Theory of Risk-Bearing* (Chicago, IL: North-Holland).

Arrow, K., 1974, Optimal Insurance and Generalized Deductibles, *Scandinavian Actuarial Journal*, 1: 1-42.

Cummins, J. D., 1993, Allocation of Capital in the Insurance Industry, *Risk Management and Insurance Review*, 3: 7-28.

Cummins, J. D., 2008, Risk-Based Premiums for Insurance Guaranty Funds, *Journal of Finance*, 43(4): 823-839.

Cummins, J. D., and O. Mahul, 2003, Optimal Insurance with Divergent Beliefs About Insurer Total Default Risk, *Journal of Risk and Uncertainty*, 27(2): 121-138.

Cummins, J. D., and O. Mahul, 2004, The Demand for Insurance With an Upper Limit on Coverage, *Journal of Risk and Insurance*, 71(2): 253-264.

Doherty, N. A., C. Laux, and A. Muermann, 2011, Insuring Non-Verifiable Losses and the Role of Intermediaries, Working paper, University of Pennsylvania.

Doherty, N. A., and H. Schlesinger, 1990, Rational Insurance Purchasing: Consideration of Contract Nonperformance, *Quarterly Journal of Economics*, 105: 243-253.

Eeckhoudt, L., and C. Gollier, 1999, The Insurance of Lower Probability Events, *Journal of Risk and Insurance*, 66(1), 17-28.

Froot, K. A. 2007, Risk Management, Capital Budgeting, and Capital Structure Policy for Insurers and Reinsurers, *Journal of Risk and Insurance*, 74: 273-299.

Froot, K. A., D. Scharsfstein, and J. Stein, 1993, Risk Management: Coordinating Corporate Investment and Financing Decisions, *Journal of Finance*, 48: 1629-1658.

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

Huberman, G., D. Mayers, and C. W. Smith, 1983, Optimal Insurance Policy Indemnity Schedules, *Bell Journal of Economics*, 14(2): 415-426.

Jaffee, D., and T. Russell, 1997, Catastrophe Insurance, Capital Markets, and Uninsurable Risks, *Journal of Risk and Insurance*, 64: 205-230.

Kaluszka, M., and A. Okolewski, 2008, An Extension of Arrow's Result on Optimal Reinsurance Contract, *Journal of Risk and Insurance*, 75(2): 275-288.

Mahul, O., and B. D. Wright, 2004, Implications of Incomplete Performance for Optimal Insurance, *Economica*, 71: 661-670.

Mahul, O., and B. D. Wright, 2007, Optimal Coverage for Incompletely Reliable Insurance, *Economic Letters*, 95: 456-461.

Merton, R. C., and A. F. Perold, 1993, Theory of Risk Capital in Financial Firms, *Journal of Applied Corporate Finance*, 6: 16-32.

Meyer, J., and M. B. Ormiston 1999, Analyzing the Demand for Deductible Insurance, *Journal of Risk and Uncertainty*, 18(2): 223-230.

Modigliani, F., and M. Miller, 1963, Corporate Income Taxes and the Cost of Capital: A Correction, *American Economic Review*, 53: 433-443.

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

Schlesinger, H., 1981, The Optimal Level of Deductibility in Insurance Contracts, *Journal of Risk and Insurance*, 48(3): 465-481.

Zanjani, G., 2002, Pricing and Capital Allocation in Catastrophe Insurance, *Journal of Financial Economics*, 65: 283-305.