Economic Impact of Capital Level in an Insurance Company

Yingjie Zhang
CNA Insurance Companies
CNA Plaza 30S, Chicago, IL 60685, USA
Tel: 312-822-4372
Email: yingjie.zhang@cna.com

Abstract

Capital level has significant impact on policy premium and shareholder return. A company with less capital is more likely to default on claim payments, so it should charge less premiums. However, the expected profit may not decline with the premium, since the expected loss payment falls as well. Shareholders provide capital and receive the insurance profit. When the deadweight costs are ignored, the shareholder return is independent of the capital level—similar to the Modigliani-Miller irrelevance theorm. With the deadweight costs, an optimal capital level exists that generates the highest shareholder return. This result is derived using a risk capital model by Perold. The no-arbitrage argument, developed by Modigliani and Miller, is used throughout the paper that puts our discussions on a solid economic footing.

Keywords

Insurance profit, shareholder return, return on capital, no-arbitrage, Modigliani-Miller irrelevance, optimal capital level.

1 Introduction

Capital level has significant impact on the policy premium and the shareholder return. Policyholders are concerned about the insurer's ability to pay claims. If the capital is inadequate, the insurer may become insolvent and default on the claims. So the economic value of the future claim payment is less than its full value. Consequently, the policyholders require a premium credit. Since the amount of credit varies with the reduction in claim payment, it is interesting to ask whether the the expected profit would change. The answer depends on the assumption. In an arbitrage-free market with no deadweight costs, the premium credit is exactly offset by the reduction in expected loss payment, so the profit remains constant, independent of the capital level and the investment return. However, when certain types of frictional costs exist, as always happens in the real world, the policyholders should require higher premium credit; hence the expected profit declines as capital decreases. Market observations support this conclusion. (Unless stated otherwise, the "market" in this paper means an integrated financial market with both investible assets and insurance policies.)

Shareholder return and risk are also affected by the capital level. Holding more capital, the company has lower risk; but the profit is divided among more shares of the stock, so the return-on-capital is also lower. Similar return/risk tradeoff is predicted by the Capital Asset Pricing Model. A natural concern is what capital level provides the best tradeoff. A related problem for non-financial firms has been solved by Modigliani and Miller [6]—the celebrated MM theorem. In an arbitrage-free world, the capital level turns out to be irrelevant. Any financial leverage the company holds can be exactly replicated by the shareholders in the capital market. We extend the MM theorem to insurance: in a default-free insurance firm, shareholders are indifferent to the capital level. If a firm is not default-free, the shareholder return and risk could be affected by the capital.

The above capital irrelevance results ignore the deadweight costs. But these costs are important in the real world. Deadweight costs in an insurance company can be divided into the frictional cost of capital and the cost of financial distress. Taxes and agency costs belong to the first type. They vary directly with the amount of capital. The second type includes costs associated with a firm's having too little capital. When a firm is threatened with insolvency, it spends extra money in dealing with policyholders, regulators, banks, employees, etc., and the profit declines. Perold [8] incorporates these costs into a "risk capital model". A change in capital level increases one type of costs and reduces the other. There is an optimal level that minimizes the total cost. The Perold model can be adapted to insurance. Since the premium varies with the capital, the adaptation needs a careful treatment of insurance profit.

Some results in this paper have been stated and discussed in actuarial literature. Our purpose is to establish a general framework in which many capital related issues can be treated with a unified and rigorous approach. We start off with an examination of basic concepts, including the market value of insured loss, the default put option and its price, and the economic balance sheet. Insurance profit, shareholder return and return-on-capital are clearly defined. The powerful no-arbitrage argument, developed in Modigliani and Miller [6], is used throughout the paper, which places our discussions and solutions on a solid economic footing. For many questions, the answer depends on the assumption. Whenever appropriate, we analyze alternative assumptions and compare results with market observations. Conclusions obtained under ideal conditions are important, like the capital independence of insurance profit. They can be formulated in simple terms and understood easily. They provide powerful insights and serve as a basis for more complex situations. On the other hand, a useful theory should possess explanatory power in the real world, where the ideal conditions usually fail. Our ultimate goal is to develop such a theory.

The rest of the paper is organized as follows. In Section 2, we begin with an examination of basic concepts, then analyze the impact of capital level and investment strategy on premium and insurance profit. Section 3 studies the shareholder return and return on capital. The MM irrelevance theorem is extended to insurance. The frictional cost of capital and the cost of financial distress are discussed in Section 4. We then apply the Perold risk capital model to show the existance of an optimal capital level.

2 Premium and Insurance Profit

Insurance profit is the gain or loss generated by issuing insurance policies. It consists of the underwriting profit and the gain from investment of premium. In this section we study the impact of the capital level on the insurance profit, under the no-arbitrage condition.

2.1 Market value of insurance losses

Insurers regard an insurance policy as an investment. After collecting the premium and paying off the claims, the insurers expect to earn a profit. So like stocks, bonds or other investment instruments, an insured loss has a *market value*, which the premium charge is based on. But the market value of insured losses are not automatically determined by the market, since the insured losses are not publicly traded. Actuaries are charged with the responsibility of calculating them—thus comes about the pricing theory. In an arbitrage-free market, every risk has a unique market value. All our results in this paper relies on this *one-price principle*. The market value is necessarily additive, since stacking or unstacking risks does not change the price of the total package.

Consider the following one-period model. The insurance policies are written at time 0. A policy loss is a random variable L and is paid at time 1. An one-year period in the real world can be approximated by the one-period model. Let l be the time-0 market value of loss L. Traditionally in insurance pricing, l is considered the present value of L plus a risk load, or the present value of L with risk adjusted discounting, see the review article Cummins, Phillips, Butsic and Derrig [2]. An asset in the one-period model is also a random variable, denoted by A. Let a be the market value of A at time 0. For a tradable asset the market value is determined by the secondary market. We introduce a *market value operator*, $V(\cdot)$, as follows. $V(\cdot)$ applies to both insured losses and assets, thus $l = V(L)$ and $a = V(A)$; and the following conditions hold

1. $V(\cdot)$ is linear with respect to both assets and insured losses;
2. $V(\cdot)$ is positive: if $A > 0$ then $V(A) > 0$, and if $L > 0$ then $V(L) > 0$;
3. $V(\cdot)$ is continuous, which roughly means the market value of a "small" asset or insured loss is also small.

These conditions are obviously satisfied by any reasonable definition of market value. A modern view of the market value is to think of it as the discounted expected value with a risk-adjusted (or risk-neutral) *probability*. In notation, $V(\cdot) = 1/(1+r) \cdot E^Q[\cdot]$, where r is the risk-free rate and E^Q the expected value with respect to a risk-neutral probability Q. This concept is originated from option pricing theory. It has gained some popularity recently. Our definition of the market value operator, V, is independent of any risk-neutral probability. But the three conditions above are inspired by this line of thought. There is no universal aggrement on the form of V. Different market models give different formulas. The articles Bühlmann [1], Phillips, Cummins and Allen [9], Wang [16] and Zhang [17] show a few examples.

2.2 The balance sheet

An insurance company raises an amount of capital c at time 0, and writes insurance policies for a total premium p. The total insurance loss is a random variable L, which will be paid at time 1. The market value of loss $l = V(L)$. In this paper the premium p is net of all expenses. The initial asset of the company is thus $a \equiv p+c$. (For a company with ongoing business, the asset is invested in a portfolio of bonds, stocks, etc. Then a is the total market value of these investments.) $k \equiv a - l$ is called the actuarial surplus. We have the following initial balance sheet

Asset	Liability
$a = p + c$	l
	k

This is the *market value balance sheet* (Phillips *et al* [9], Myers and Read [7]). It is usually different from the statutory or the GAAP balance sheet. Since the statutory and the GAPP values are subject to arbitrary accounting rules, the market values are more relevant in pricing and risk management.

In the market value balance sheet, the liability is valued as if it can be 100% paid at time 1. In fact, when a company becomes insolvent, the claiments cannot recover their losses in full. So if there is a chance of insolvency, the true value of liability is less than the market value. Consequently, the true value of the policyholder surplus is greater than the actuarial surplus k. The balance sheet may be restated to show these effects explicitly.

The asset is invested to earn an income. Let the average rate of return be R. The time 1 value of asset is $A = a(1 + R) = (p + c)(1 + R)$. The company becomes insolvent if $L > A$. Upon insolvency the insurer pays out the entire asset A to the policyholders. The shareholders assume no more liability. The amount of policyholder unrecoverable loss is $D \equiv \max(L - A, 0)$. D is usually considered a put option on the company asset owned by the shareholders, exercised upon default. The price of the default put, $d \equiv V(D)$, is also known in the actuarial literature as the expected value of default, or the expected policyholder deficit.

Clearly, the time 1 actual payoff is $L - D$. So the true value of liability is $l - d$, and the true policyholder surplus is $a - (l - d) = k + d$. We have the following *economic balance sheet*

Asset	Liability
$a = p + c$	$l - d$
	$k + d$

This "more precise" balance sheet has appeared in, among others, Merton and Perold [5], Myers and Read [7] and Sherris [11]. The economic capital, $k + d$, is the market value of the company net asset, after the claims are paid off.

2.3 Fair premium and insurance profit

The fair premium of an insurance policy is the market value of the covered loss (ignoring expenses and capital costs). If a company is default-free, the fair premium equals $l = V(L)$. If it is not default-free, then the fair premium is reduced. As discussed previously, the policyholders receive claim payment $L - D$ at time 1. So the fair premium is its market value $l - d$. This is a basic principle of risk premium theory (Phillips *et al* [9], Myers and Read [7]).

If the fair premiums are charged, the market value of the insurance profit is zero. The company breaks even. In an equilibrium state, all companies expects zero profit. The real insurance market, however, never stays in equilibrium. The premiums and profits vary in a cyclical pattern—the underwriting cycle. The forces underlying this pattern are numerous and complex. For Property/Casualty insurance, a thorough analysis of underwriting cycle can be found in Feldblum [3]. Although premiums vary from time to time, all companies seem to move in-sync in raising or reducing the premium for a given line of business. Thus at any point in time, the one-price principle is still a good approximation. We discuss the implication of the one-price principle in a market that is not necessarily in equilibrium.

Let p be the total premium actually charged by a company. For a given investment return R (a random variable), the premium grows to $p(1 + R)$. The claim payoff is $L - D$. The *insurance profit* is thus

$$Y \equiv p(1 + R) - (L - D). \tag{2.1}$$

Traditionally in actuarial literature, $p - (L - D)$ is called the underwriting profit, and pR the investment income. Their sum, the insurance profit, is the total income generated by issuing policies. The market value of insurance profit is

$$y \equiv V(Y) = p - (l - d). \tag{2.2}$$

A given block of business L may be insured by different companies, with different capital level and investment strategy. If the company is default-free, then the policyholders receive full claim payment L. By the one-price principle, all default-free companies should charge the same amount of premium, denoted by p^0. This premium is determined by the current phase of the market cycle, usually different from the fair premium. The profit earned by a default-free company is $p^0(1 + R) - L$, with market value $y^0 = p^0 - l$, usually not zero. The profit margin may vary across market segments, or from policy to policy. Although companies would like to do business in the more profitable segments, they sometimes issue policies with less or even negative expected profit for the purpose of retaining desired market shares. Ideally, the negative profits only exist in the short run.

If a company is not default-free, the capital level and the investment return affects the default put D, as well as its market value $d = V(D)$. The premium should be reduced according to the size of the expected default. We have seen if p^0 is the fair premium charged by a default-free company, then $p^0 - d$ is the fair premium for a company that is not default-free. We will explore similar relationship in the situation that p^0 is not necessarily fair.

2.4 The solvency guarantee

A company may purchase insurance to cover its default risk. Such an insurance is called a solvency guarantee. A company with a solvency guarantee is practically default-free. Merton and Perold [5] studied this concept extensively. They use the term "risk capital" for the fair price of the guarantee. The guarantee does not have to be actually purchased in an insurance transaction. It may take the form of parent guarantee or the residual market guarantee. Regardless of which party serves as the guarantor, the same fair price should be paid by the company. If the guarantor is the residual market, the company pays an assessment fee. To be fair to all residual market participants, the assessment should be risk based and equal the fair price. In the case of the parent guarantee, a portion of the parent firm capital is allocated for this purpose, whose amount again should equal the fair price. In the following discussion, without loss of generality, we can think of the guarantee as actually purchased.

A solvency guarantee completely eliminates the default risk. The guarantor pays an amount $L - A$ at time 1 when $L - A > 0$. The guarantor's liability is exactly the policyholders' unrecoverable loss $D = \max(L - A, 0)$. Its market value is $d = V(D)$. Ignoring the transaction expense, this is the fair premium the company should pay for the guarantee. d is inversely related to the capital level: the lower the capital, the more solvency guarantee is needed, thus the larger d. In a sense, the solvency guarantee is a substitute for capital. Because the distribution of liability L usually has a long thin tail on the side of large losses, a small amount of additional d can trade for a large reduction in capital. Companies can also buy a partial guarantee that reduces, but not eliminates, the default risk.

Most insurance companies operate with some default risk. This is surely the case without a solvency guarantee. Even with the guaranty of a residual market or a parent, there is still some risk of unrecoverable claims. Recoveries from the residual market may be delayed and limited. The parent firm may have its own financial trouble. Merton and Perold [5] observes that issuing policies with default risk is equivalent to purchasing a solvency guarantee from the policyholders. The price of the guarantee partially offsets the insurance premium.

2.5 Invariance of insurance profit

If a company raises (reduces) its capital level, it assumes more (less) liability when writing the same policy. At the same time it also charges more (less) premium. The amount of liability assumed and the corresponding premium are also affected by the investment return and risk. Thus the market value of profit, y, varies with capital level c and investment return R. Our first result is that, under ideal conditions, y is actually constant.

Given a set of policies with total random loss L. Assume the market allows a default-free company to charge amount of premium p^0. p^0 may not equal the fair premium. The one-price principle implies every default-free company should charge the same premium p^0. Now consider a company that is not default-free. It holds initial capital c. It charges premium p for the policies, and invests the asset $p+c$ with random return R. p is p^0 less a deduction. The market value of default is $d = V(D) = V(\max(L-(p+c)(1+R), 0))$. The company can become default-free by purchasing a solvency guarantee. The fair premium for the guarantee is d. We can prove d is also the premium reduction.

Proposition 1 *If solvency guarantees can be purchased with the fair premium, then* $p = p^0 - d$.

In fact, the policyholders can split the liability L as follows. They pay a premium p to insure the limited loss $\min(L, (p+c)(1+R))$, and pay a separate premium to a default-free guarantor to cover the residual liability $\max(L - (p + c)(1 + R), 0)$. By assumption, the second coverage can be purchased with fair premium d. The combined coverage is default-free, so should be priced at p^0. By the one-price principle, we have $p + d = p^0$, proving the proposition.

If the proposition holds, d is a function of the capital level c, the default-free premium p^0, and the investment return R. It is solved from the following equation

$$d = V\left(\max(L - (p^0 - d + c)(1 + R), 0)\right). \tag{2.3}$$

Normally the equation has a unique solution between 0 and p^0. The market value of the insurance profit is $y = p - (l - d)$. For a company with no default risk, it is $y^0 = p^0 - l$. Applying Proposition 1 we have $y = y^0$.

Corollary 1 *Under the condition of Proposition 1, the market value of the insurance profit is independent of the capital level and the investment strategy.*

Thus the insurance profit of writing a risk is determined by the market conditions at the time, no matter which company issues the policy. (The profit margin usually differs from policy to policy.) As a special case, if the market allows exactly the fair premium, as in the market equilibrium, then the market value of profit is zero regardless of the likelihood of default. In market equilibrium, all policies are charged with the fair premium (Phillips *et al* [9], Myers and Read [7], Cummins *et al* [2]).

The key assumption in Proposition 1 and the invariance of profit is that solvency guarantees can be purchased with their fair premium d. This is theoretically true for the various forms of guarantee mentioned in Section 2.4. The guaranty fund is set up to make the insurance market functional, not to earn a profit. So companies should be assessed in exactly such amount that the fund breaks even. The case of parent guarantee is similar: the parent firm should hold an amount d for the guarantee. However, we have argued that the guaranty funds or the parent firms seldom provide full guarantee in the real world. There is usually some risk of unrecoverable claims borne by the policyholders. In other words, the insurer buys at least part of the solvency guarantee from the policyholders. If this happens, the policyholders incur more costs than simply taking back the unrecovered claims. They may have to pay expenses on arbitrations and lawsuits, and experience delay in payments.[1] Therefore, the price of a full solvency guarantee is higher than its fair price. The one-price principle then implies $p = p^0 -$ price of full solvency guarantee $< p^0 - d$, meaning that the policyholders demand a premium credit greater than d. Thus the insurance profit, rather than invariant with capital level, is reduced when the firm is not default-free.

Empirical studies in Sommer [12] and Phillips *et al* [9] confirms that premium declines as the possibility of default increases. See also Froot *et al* [4] for a survey of literature. Further, Wakker, Thaler and Tversky [15] and Phillips *et al* [9] show that premium credits requested by the policyholders are often many times more than the expected value of default. Wakker *et al* [15] cites psychological reasons and the prospect theory to explain the findings. We have provided another explanation using the one-price principle.

Insurance profit determines the shareholder return. As the likelihood of insolvency increases, the premium declines. But as long as the insurance profit stays unchanged, the shareholders will earn the same income. The impact of capital level and investment return on the shareholders is studied in the next section.

3 Shareholder Return on Capital

The Modigliani-Miller theory (MM) is a cornerstone of modern finance. MM's Proposition I (MMI) states that, if taxes are ignored, the value of a firm is independent of its capital structure (Modigliani and Miller [6]). The MM article develops a "no-arbitrage" argument, which is now one of the most celebrated techniques in finance. There has been a great deal of literature about the MM. A modern view of the theory and review of historical development can be found in Rubinstein [10]. The MM irrelevance result holds under the conditions of no tax and no frictional costs. It is a "first-order effect". Although only dealing with ideal situations, the first-order effects are nonetheless important. The results can be formulated clearly and easily understood. They provide powerful insights and serve as a basis for analyzing more complex real wold situations. Higher-order

[1] I thank Gary Venter for providing this argument.

effects can be studied as deviations from first-order ones.

To extend MM to the insurance world, we may view the insurance policies as a type of risky debt instrument. The liability to capital ratio resembles the financial leverage. However, the MM is not directly applicable in this case, since the riskiness of the debt invalidates the no-arbitrage argument. The business type also makes a difference. The original MM addresses the financing issue of product firms, like manufacturers and retailers, whose main activity is on the asset side of the balance sheet. The primary business of an insurance firm, however, is issuing insurance policies. The shareholder risk and return are dominated by the liability side of the balance sheet. Thus the irrelevance results take on a different form. Vaughn [13] discusses an insurance MM in a heuristic fashion.

3.1 Capital irrelevance in default-free firms

Consider an insurance company with equity financing but no debt. Shareholders contribute capital at the beginning of the policy term, and receive the remaining asset at the end, after claims are paid off. Assume the company has a set strategy to invest the premium. Regardless of capital and premium level, the premium is always invested in a stock and bond portfolio with a fixed percentage for each security. Let R be the rate of return of the portfolio. Also assume the company will not default. Then the insurance profit is $Y = p(1+R)-L$. Since the market value of profit, $y = p - l$, is independent of how the premium is invested, it is sensible to select an investment strategy to maximize the expected value and minimize the risk of Y. (In the model of Perold [8], $p(1 + R)$ is a hedging portfolio for L and Y the hedging error. But hedging the liability is less viable in Property/Casualty insurance.) We examine the effect of capital level in two situations, when the capital is invested with or without risk, respectively. We prove irrelevance results similar to the MM.

Consider two insurance companies with different capital levels that are identical otherwise. Company #1 starts with capital amount c^1 and insures liability L^1. Company #2 with capital c^2 and liability L^2. $c^1 > c^2$. L^1 and L^2 are identically distributed and perfectly correlated random variables. Thus L^1 and L^2 have the same market value, $l \equiv V(L^1) = V(L^2)$. Assume c^1 and c^2 are so large that both companies are default-free. From Section 2, the companies charge the same market determined premium $p \equiv p^1 = p^2$. The premiums are invested in the same stock and bond portfolio with risky return R. Assume the capital of each company is invested risk-free. The insurance profit of each company is $Y^j = p(1+R) - L^j$, which are also identically distributed and perfectly correlated, with the same market value $y = V(Y^1) = V(Y^2) = p - l$.

Suppose an investor holds an amount of cash c^1. He can either invest the entire c^1 in company #1, or invest an amount c^2 in company #2 and the rest, $c^1 - c^2$, in the risk-free asset. In the second way, at time 1, the investor receives $p(1+R) + c^2(1+r) - L^2 = Y^2 + c^2(1+r)$ from company #2, and $(c^1 - c^2)(1+r)$ from the risk-free asset. The total cash is $Y^2 + c^1(1+r)$, which is identically distributed and perfectly correlated with $Y^1 + c^1(1+r)$, the net asset from investing in #1. Therefore, it makes no difference which company he invests in.

If the investor holds an amount c^2, instead. He can either invest c^2 in company #2, or borrow an amount $c^1 - c^2$ risk-free, and invest the sum, $c^1 = c^2 + (c^1 - c^2)$, in #1. Once again, either way he receives the same amount of cash at the end. We have the following conclusion.

Proposition 2 *If a firm is default-free and the capital is invested risk-free, an investor is indifferent to the capital level.*

This may be considered a version of the MM for the insurance company. One condition for the theorem to hold is the investor can borrow and lend at the risk-free rate. Usually it costs individuals more to borrow than to lend. So investors would not choose to borrow in order to contribute more capital. This would make the lower capital level more attractive.

Instead of keeping the capital in a risk-free account, a more realistic assumption is to invest it just like the premium. A weaker form of irrelevance result exists in this situation. If both the premium and the capital grow at the random rate R, the asset at time 1 is $(p+c)(1+R)$. Again look at the two companies #1 and #2, and use the same notations as before. Shareholders can construct replicating investments similarly. An investor holding cash c^1 can invest an amount c^2 in firm #2 and the rest, $c^1 - c^2$, in the same portfolio as the asset, earning return R. The combined investment replicates the return of firm #1. On the other hand, if the investor holds cash c^2, he can short sell the portfolio with return R for amount $c^1 - c^2$, invest the proceeds plus the initial c^2 in firm #1, thus replicates the return of firm #2. This proves the value of the firm remains the same to the investors, regardless of the capital level. But if short sale is expensive, then the firm with less capital becomes more valuable.

In the first situation where the capital is risk-free, the two firms differ by a risk-free asset, just like the original MM. In the second situation, they differ by some risky asset in the capital market. It is not an irrelevance result in the sense of MM. But still, since the shareholders have free access to the capital market, they do not care about the difference between the two firms. We have a weaker form of capital irrelevance. As already seen, the profit Y^j varies with the investment return R. The investment strategy should target at optimizing the return/risk tradeoff of Y^j.

3.2 Impact of capital level when default is possible

The above irrelevance results hold in default-free firms, as in the original MM. Capital level does matter when the chance of default is great than zero. Because of the limited liability, the shareholders shift the risk of default to the policyholders. This risk can not be replicated in the capital market.

Reexamine the two-company setup in Section 3.1. Now assume company #2 may become insolvent, but #1 stays default-free. The assets are invested with random rate of return R. In company #2 the premium is lower according to the value of default. The default put of the company is $D^2 = \max(L^2 - (p^2 + c^2)(1 + R), 0)$, and its market value $V(D^2) = d^2$. Assume Proposition 1 holds. Then $p^2 = p^1 - d^2$. The companies make insurance profit $Y^1 = p^1(1 + R) - L^1$ and $Y^2 = p^2(1 + R) - (L^2 - D^2)$, respectively. They have the same market value $y = V(Y^1) = V(Y^2) = p^1 - l$.

An investor with cash c^1 may invest in company #1. He receives net asset

$$S^1 \equiv (p^1 + c^1)(1 + R) - L^1 = Y^1 + c^1(1 + R) \tag{3.1}$$

at time 1. Or he may choose to invest an amount c^2 in company #2 and the remaining amount $c^1 - c^2$ in the same portfolio as the asset. At time 1 his net asset in company #2 is

$$S^2 \equiv (p^2 + c^2)(1 + R) - (L^2 - D^2) = Y^2 + c^2(1 + R). \tag{3.2}$$

So his combined net asset is

$$S^2 + (c^1 - c^2)(1 + R) = (p^1 + c^1)(1 + R) - L^2 + D^2 - d^2(1 + R). \tag{3.3}$$

Compared with (3.1), this investment scheme does not replicate S^1 exactly. The error term is $D^2 - d^2(1+R)$. In general the default D^2 can not be hedged in the capital market. Therefore, when the default possibility exists, the capital level is relevant. The shareholders of company #2 hold the default put option, D^2, but they paid for it at the fair market price d^2.

Since the market value of $D^2 - d^2(1+R)$ is 0, the market value of the combined investment (3.3) equals $V(S^1)$. This shows another way of looking at the capital irrelevance: by contributing a different amount of capital, the shareholders hold a different combined investment; But all these investments have the same market value. This is not the case, however, when Proposition 1 does not hold. As discussed in Section 2.5, the premium reduction in company #2 is usually more than d^2. Thus the combined investment (3.3) has lower market value than $V(S^1)$, meaning the default-free company is more appealing to shareholders. The greater the default risk, the less valuable the company is.

3.3 Return on capital

MM's Proposition II (MMII) establishes a relationship between the rates of return on various investments

$$R_e = R_a + \frac{d}{e}\,(R_a - R_d), \tag{3.4}$$

where R_a is the rate of return on asset, R_e the rate of return on shareholder equity, and R_d the interest rate to debtholders. d and e are the initial value of debt and equity, and d/e the "leverage ratio" of the firm. We want to derive similar equations for insurance firms.

Assume a company is default-free and the asset is invested with risky return R. The end-of-period net asset is given by

$$S = (p+c)(1+R) - L = Y + c(1+R). \tag{3.5}$$

The rate of return on capital is

$$\text{ROC} \equiv \frac{S-c}{c} = R + \frac{p}{c}\cdot\left(R - \frac{L-p}{p}\right) \tag{3.6}$$

$$= R + \frac{Y}{c}. \tag{3.7}$$

We can think of $(L-p)/p$ as the rate of retun on premium for policyholders, and p/c the leverage ratio. Then equation (3.6) is parallel to the MMII equation (3.4). (Vaughn [13] defines the leverage ratio as—using our notation—l/k. It only works in the equilibrium situation, where $p = l$, and $c = k$.)

Write (3.7) as

$$\text{ROC} - R = Y/c. \tag{3.8}$$

The left-hand side is the rate of return on capital in excess of the asset rate of return. It varies in inverse proportion to the capital amount c, as Y is independent of c . However, this does not mean the shareholders are better off with less capital in the firm. In fact, we have shown in Section 3.1 that the capital level is irrelevant in a default-free firm. Here is another way to look at (3.8). The Capital Asset Pricing Model states that, for fairly priced stocks, the expected excess rate of return is proportional to its β. β is a measure of the stock's systematic risk, and is in direct proportion to the covariance between the stock's return and the market return. Thus, investors expect higher returns from more risky stocks, and lower returns from less risky ones. Trading one for another with fair market prices does not make an investor better or worse off. In (3.8), although reducing c by half doubles the excess rate of return, it also doubles the risk, measured by either the standard deviation of Y/c, or the covariance of Y/c with some market portfolio. So it does not generate more wealth for the shareholders. Actually, the irrelevance result in Section 3.1 tells a little bit more. The investors of firms with different capital levels can replicate each other's return by investing in or short selling the same portfolio as the asset.

(3.8) can be written as $c \cdot (\text{ROC} - R) = Y$. This says the total excess return to all shareholders equals the insurance profit, regardless of the capital amount. Investors can assemble a stock and bond portfolio themselves to generate return R. They are attracted to the insurance firm because they expect a share of insurance profit, Y, in excess of the portfolio's return. Another possible reason of investing in an insurance firm is for diversification, since Y is largely independent of the capital market movement.

Now look at a more typical company that is not default-free. As in (3.2) the net asset at time 1 is

$$S = (p + c)(1 + R) - L + D = Y + c(1 + R), \tag{3.9}$$

where p is the reduced premium as in Theorem 1. The rate of return on capital is

$$\text{ROC} = \frac{S-c}{c} = R + \frac{p}{c} \cdot \left(R - \frac{L-D-p}{p} \right) \tag{3.10}$$

$$= R + \frac{Y}{c}. \tag{3.11}$$

$(L-D-p)/p$ is the rate of retun on premium for policyholders, and p/c the leverage ratio—same as in the default-free case. Equation (3.10) is again a version of the MMII for insurance firms. But when the firm is not default-free, we do not have a MMI type of capital irrelevance result. (3.11) can be written as $c \cdot (\text{ROC} - R) = Y$, again meaning the total excess return to the policyholders equals the insurance profit. Although the insurance profit, Y, varies with the capital level, its market value may stay constant. The market value of the total excess return is

$$c \cdot V(\text{ROC} - R) = y. \tag{3.12}$$

This equation applies whether the firm is default-free or not. y is a constant if the firm is default-free, or if Proposition 1 holds. In these cases the market value of the total excess return is independent of the capital level and the investment return— again showing the only way to increase shareholder wealth is to make a positive insurance profit y, i.e., to charge more premium. If Proposition 1 does not hold, as discussed in Section 2.5, then y declines as the likelihood of default increases. Thus firms with more adequate capital are more valuable.

The last statement means that financially weak firms are not worth investing. When insuring similar lines of business, the weak firms generate less shareholder return than the strong ones. Thus the weak firms need to differentiate themselves or to grow stronger in order to survive. One solution is to move to specialized markets to avoid the competition. Another is to inject more capital into the firm. Doing so not only better satisfies the regulatory requirements, but also increases the insurance profit and the shareholder return. Adopting risk management measures, such as reinsurance or diversification, may also be helpful.

4 Deadweight Costs and Optimal Capital Level

The previous sections focus on the first-order effects. Taxes and other deadweight costs are ignored. In real world, however, these costs are significant. The deadweight costs can be roughly divided into two categories. The frictional costs of capital vary in proportion to the amount of capital, while the costs of financial distress increase with the likelihood of insolvency. Perold [8] develops a model of financial firms and studies the impact of these costs. There is an optimal capital level that minimizes the total deadweight costs. We adapt the Perold model to insurance firms.

4.1 Frictional costs of capital

Frictional costs of capital are similar to the friction in classical mechanics. If we define an insurance system as the collection of policyholders and shareholders, joined together by the insurance policies, then the frictional cost is a wealth transfer from the insurance system to outside agents. From the point of view of policyholders and shareholders, the wealth disappears like heat dissipation. (Underwriting and other expenses are also such wealth transfers. But only costs beyond the "normal" category or range are considered frictional costs.) Examples of frictional costs include the double taxation and agency costs. The investment income of capital is first taxed at the corporate level and then taxed again when shareholders take the realized capital gain. This double taxation creates a cash flow from the insurance system to the government. It a direct cost to the shareholders, and indirectly affects the policyholders through increased premium. The agency cost is also significant. As pointed out in Merton and Perold ([5]), a distinguishing feature of the financial firm is its "opaqueness" to the customers and shareholders. As a consequence a financial firm usually experiences high agency costs in utilizing the capital. The agency costs transfer wealth from the policyholders and the shareholders to the firm managers and employees, internal or external. Insurance regulations also contribute some frictional costs (Venter [14]).

The frictional costs of capital should not be confused with the cost of capital. The latter is a widely used term in corporate finance. It is a cost to the *firm* to finance projects. For a stock company, it means the shareholder required rate of return. In the paradigm of the Capital Asset Pricing Model, the required rate of return varies directly with the systematic risk β. Unlike the frictional costs, the cost of capital is not a loss to the insurance system, because it increases the shareholder wealth.

In Perold [8], the total frictional cost is assumed to be proportional to the net asset of the firm at time 1. In our notation it is δS, where δ is a constant and the net asset S is given in (3.9). The multiplier δ's being a constant seems appropriate under normal operating conditions. If the business pattern changes drastically, the cost could be a non-linear function of the net asset. Our discussion is limited to the Perold scenario.

4.2 Costs of financial distress

As the capital amount decreases, the frictional cost is reduced. However, a lower capital level means a higher chance of financial distress. This drives up the other type of deadweight costs—the costs of financial distress. Perold [8] calls it the "monitoring charge". When a firm buys the solvency guarantee from an external guarantor, the guarantor monitors the firm to protect itself against adverse selection and moral hazard. It charges a fee for this purpose. Clearly, this charge increases with the likelihood of insolvency. Similar charges exist even without an external guarantor. Froot *et al* [4] lists many cost items associated with financial distress. Extra expenses are incurred in dealing with auditors and regulators, and in defending lawsuits. As the firm's opportunity set shrinks and employee morale deteriorates, the productivity declines. Just as the firm needs external fundings the most, they become increasingly expensive to obtain. The total cost of financial distress is a decreasing function of capital level and an increasing function of the likelihood of default.

Following Perold [8], we assumes the cost of financial distress is of the form μd, where d is the value of default put and the multiplier μ is a constant. The linear form simplifies analysis. When the capital level becomes extremely low, the cost may increase more quickly than linear. However, for our goal of seeking the optimal capital level, the linear function should be a reasonable approximation.

4.3 The optimal capital level

The deadweight costs serve to reduce the shareholder return. Thus additional amount of premium should be charged to cover these costs. On the other hand, the total deadweight costs can be minimized with the right amount of capital. Without deadweight costs, the shareholders receive S at time 1. Their excess return is $S - c(1+R) = Y$ (equation (3.9)), and $V(Y) = y$. The total deadweight cost is the sum of the frictional cost and the cost of financial distress. Its market value is $\delta \cdot V(S) + \mu d$. Therefore, the shareholder excess return net of the deadweight cost has the following market value

$$y - \delta \cdot V(S) - \mu d. \tag{4.1}$$

Under the condition of Proposition 1, y is a constant. To maximize the excess return (4.1) is to minimize the market value of the deadweight cost

$$\delta \cdot V(S) + \mu d. \tag{4.2}$$

Thus the optimal capital amount c minimizes equation (4.2), for given constants δ and μ. We have seen in Section 3 that without the deadweight costs, i.e., $\delta = \mu = 0$, the capital amount is irrelevant, and the shareholder excess return equals the insurance profit. In the real world, where δ and μ are both positive, shareholders should contribute the optimal amount of capital to the company and invest the rest of their assets elsewhere.

The following result is expected.

Proposition 3 *There is a unique capital level c that minimizes the market value of deadweight cost (4.2).*

We have extended Perold [8] in several ways. First, insurance has a unique feature that the capital level affects premium. Hence the firm asset is related to the capital in a more complex way. This case was not treated in Perold [8]. Second, the capital is invested in risky securities—a more realistic scenario than Perold [8]. Third, our result holds for any distribution type of L and R, and does not depend on option pricing. Although used extensively in theoretical studies, the option pricing theory rests on some restrictive assumptions whose applicability to insurance liabilities has not been confirmed.

In Appendix at the end of the paper, we provide formulas for the optimal capital. The ratio μ/δ determines the best tradeoff between the frictional cost and the cost of financial distress. The larger the ratio, the more costly the financial distress relative to normal frictions; thus the higher the optimal capital. There are evidences that μ is usually many times larger than δ. Further research is needed to measure the two constants. In Perold [8], the minimum deadweight cost was expressed as a function of δ and μ times the standard deviation of the profit. But that result highly depends on the option pricing formula he used and can not be properly generalized.

If Proposition 1 fails, the shareholder return before the deadweight costs declines as the capital level falls. This effectively increases the cost of financial distress. Therefore, the optimal capital level is higher than what is obtained in

Proposition 3.

Shareholders may require a company to be run with the optimal level of capital. It is the management's responsibility to study the company business and the economic environment, and determine the optimal level. Many actions could be taken to reduce the deadweight costs and utilize the capital more efficiently.

References

[1] Bühlmann, H. (1980) An Economic Premium Principle, *ASTIN Bulletin*, **11(1)**, 52-60.

[2] Cummins, J.D., Phillips, R., Butsic, R. and Derrig, R. (2000) The Risk Premium Project (RPP) Phase I and II Report, *Casualty Actuarial Society Forum*, Fall 2000, 165-230.

[3] Feldblum, S. (2001) Underwriting Cycles and Business Strategies, *Proceedings of the Casualty Actuarial Society*, LXXXVIII, 175-235.

[4] Froot, K., Venter, G. and Major, J. (2004) Capital and Value of Risk Transfer, *14th Annual International AFIR Colloquium*, 181-195.

[5] Merton, R. and Perold, A. (1993) Theory of Risk Capital in Financial Firms, *Journal of Applied Corporate Finance*, **6** (Fall), 16-32.

[6] Modigliani, F. and Miller, M. (1958) The Cost of Capital, Croporation Finance and the Theory of Investment, *American Economic Review*, **48(3)**, 261-297.

[7] Myers, S. and Read, J. (2001) Capital Allocation for Insurance Companies, *Journal of Risk and Insurance*, **68(4)**, 545-580.

[8] Perold, A. (2005) Capital Allocation in Financial Firms, *Journal of Applied Corporate Finance*, **17(3)** (Summer), 110-118.

[9] Phillips, R., Cummins, J.D. and Allen, F. (1998) Financial Pricing of Insurance in the Multi-Line Insurance Company, *Journal of Risk and Insurance*, **65(4)**, 597-636.

[10] Rubinstein, M. (2003) Great Moments in Financial Economics: II. Modigliani-Miller Theorem, *Journal of Investment Management*, **1(2)**, 7-13.

[11] Sherris, M. (2004) Solvency, Capital Allocation and Fair Rate of Return in Insurance, preprint. Online at http://www.casact.org/research/summaries/allocation.htm.

[12] Sommer, D. (1996) The Impact of Firm Risk on Property-Liability Insurance Prices, *Journal of Risk and Insurance*, **63(3)**, 501-514.

[13] Vaughn, T. (1998) The Impact of Investment Strategy on the Market Value and Pricing Decisions of a Property/Casualty Insurer, *Proceedings of the Casualty Actuarial Society*, LXXXV, 211-244.

[14] Venter, G. (2003) Capital Allocation: An Opinionated Survey, *Casualty Actuarial Society Forum*, Summer 2003, 279-305.

[15] Wakker, P., Thaler, R., and Tversky, A. (1997) Probabilistic Insurance, *Journal of Risk and Uncertainty*, **15**, 7-28.

[16] Wang, S. (2003) Equilibrium Pricing Transforms: New Results Using Bühlmann's 1980 Economic Model, *ASTIN Bulletin*, **33(1)**, 55-73.

[17] Zhang, Y. (2006) A Portfolio Theory of Market Risk Load, *Casualty Actuarial Society Forum*, Winter 2006, 503-530.

Appendix: The Optimal Capital Level

In general there is no explicit formula for either the optimal level of capital or the corresponding optimal shareholder net return. The following discussion may provide insights and assist in calculation. Define the set of default events—the default set—as

$$\Omega \equiv \{\omega | A - L < 0\} = \{\omega | a(1 + R) - L < 0\}. \tag{4.3}$$

Let 1_Ω be the characteristic function of Ω: $1_\Omega = 1$ on the set Ω and $1_\Omega = 0$ elsewhere. It can be proved that the optimal default set Ω^* satisfies

$$V((1 + R) \cdot 1_{\Omega^*}) = \frac{\delta}{\delta + \mu}. \tag{4.4}$$

This equation determines the unique optimal asset a^*. The corresponding optimal capital c^* is given by

$$c^* = a^* - p^0 + V(\max(L - a^*(1 + R), 0)), \tag{4.5}$$

where p^0 is the default-free premium. (4.5) is derived from equation (2.3). If the asset is invested to earn the risk-free return r, then equation (4.4) reduces to

$$(1 + r)V(1_{\Omega^*}) = \frac{\delta}{\delta + \mu}. \tag{4.6}$$

If there is a risk neutral probability Q, so that $V(\cdot) = 1/(1 + r) \cdot E^Q[\cdot]$, then (4.6) can be further simplified

$$Q[\Omega^*] = \frac{\delta}{\delta + \mu}. \tag{4.7}$$

That is, the optimal capital level is such that the probability of insolvency equals $\delta/(\delta + \mu)$. Perold [8] uses (4.7) to derive an expression for the optimal shareholder net return. But that result cannot be generalized.

Equation (4.5) can be combined with either (4.4) or (4.6) or (4.7) to practically calculate the optimal capital. Obviously, the optimal capital is a function of ratio μ/δ. The larger the ratio, the more costly the financial distress relative to the normal friction; therefore the higher the optimal capital.