Volume Title: Studies in Public Regulation

Volume Author/Editor: Gary Fromm, ed.

Volume Publisher: The MIT Press

Volume ISBN: 0-262-06074-4

Volume URL: http://www.nber.org/books/from81-1

Publication Date: 1981

Chapter Title: Theory of Solvency Regulation in the Property and Casualty Insurance Industry

Chapter Author: Patricia Munch, Dennis Smallwood

Chapter URL: http://www.nber.org/chapters/c11431

Chapter pages in book: (p. 119 - 180)

3

Theory of Solvency Regulation in the Property and Casualty Insurance Industry

Patricia Munch
Dennis Smallwood

The objective of this article is to examine the case for solvency regulation of the property and casualty insurance industry and to examine the effects of regulation in its current form. The case for solvency regulation clearly derives from the difficulty of a policyholder in establishing the financial soundness of alternative firms. But policyholders are not the only parties concerned about the possibility of insolvency. A firm's owners also lose; in fact, they lose their equity completely, whereas policyholders and claimants may receive partial coverage. The insolvency risk is not determined exogenously, but is a byproduct of conscious choices taken to advance owner objectives. This study is motivated by the question: Under what conditions are the interests of owners sufficient to provide policyholders with an adequate level of protection?

Previous analyses have focused on two aspects of the problem: the intrinsic risks in writing insurance, and managerial incompetence or dishonesty. Analyses that focus on the intrinsic risks of the insurance business concentrate on the statistical properties of the loss distributions and on statistical-ruin problems. Taking the parameters of the loss distributions as given, such studies attempt to determine the level of surplus or reserves necessary to reduce the probability of insolvency to some small arbitrarily chosen level. The implicit (sometimes explicit) assumption is that insolvencies occur only because regulators have not applied the sophisticated mathematical and computational tools that are necessary to establish required capital requirements (see, for example, Hammond et al. 1978; Hofflander 1969). It is confidently hoped that improvements in regulatory skills can and should reduce or eliminate the insolvency problem.

Those who stress the importance of managerial incompetence or dishonesty (see McKinsey and Co. 1974) pin their hopes on more frequent examinations, and better trained auditors. In this case, there is an implicit assumption that a clear distinction exists between the behavior of firms that become insolvent and honest, well-managed firms. We ignore problems of fraud and dishonesty not because we deny their existence, but because they constitute a different problem that is amenable to a different type of analysis and regulatory response.[1]

We adopt the position that both the underlying risks of the insurance business and the behavior of management are important. We assume that managers will accept those risks that maximize the value of the firm to its owners. Thus, the underlying statistical properties of claims distributions and investment-returns distributions are relevant. But we reject the presumption that the risks accepted by the firm are exogenous. Rather, we assume that the risk of insolvency is selected by a management that is competent but is not motivated to avoid all risks, at any cost.

The article thus focuses on the choices of the firm that implicitly determine the probability of insolvency. An analysis of regulations to reduce the likelihood of insolvencies must consider how such regulations affect managerial decisions.

One-Period Model of an Insurance Firm

In this model, at the beginning of a given period the owners of a firm provide financial "capital" equal to K. The firm then sets a premium rate P and underwrites Q policyholders.[2] For simplicity, all claims and other costs are incurred at the end of the period. The ith policyholder imposes a cost C_i on the firm, which represents both the total claims cost for the ith policyholder (including both claims payments and "loss adjustment expenses") and the costs of writing and administering the policy. Thus, C_i has a positive, nonrandom component, but we shall nevertheless refer to C_i as the claims cost of the ith policyholder for convenience. Fixed, overhead insurer costs are ignored. Thus, total insurer costs, T, are

$$T = \sum_{i=1}^{Q} C_i \tag{1}$$

and we let $\bar{C}$ denote average realized claims cost per policyholder:

$$\bar{C} = T/Q = \left(\sum_{i=1}^{Q} C_i\right)\Big/ Q \tag{2}$$

so that $T = \bar{C}Q$.

Policyholders are assumed to have identical claims probability distributions and to impose equal nonrandom costs on the insurer. Thus, expected claims cost, denoted E_C, and the variance in claims cost, denoted σ_C, are equal for all policyholders:

$$E(C_i) = E(C_j) = E_C \qquad \text{for all } i,j, \tag{3}$$

$$\mathrm{Var}(C_i) = \mathrm{Var}(C_j) = \sigma_C = (S_C)^2 \qquad \text{for all } i,j. \tag{4}$$

Claims costs of different policyholders are not assumed to be independent random variables. The pairwise correlation coefficient for any two policyholders is assumed equal and is denoted by γ:

$$\mathrm{Cov}[C_i, C_j] = \gamma\sigma_C \qquad \text{for all } i,j. \tag{5}$$

The assumption that claims costs of different policyholders are not independent implies that the variance per policyholder on "underwriting" does not necessarily approach zero as the number of policyholders increases, as is usually assumed. We demonstrate in appendix C (at the end of the article) that this aspect of the model can be interpreted as reflecting the firm's "uncertainty about the distribution of claims." That is, allowing $\gamma \neq 0$ widens the range of interpretation of the model to include the case where the firm is uncertain about the parameters of the claims distribution. While the firm may be expected to eventually infer the parameters of a stable claims distribution, uncertainty may persist if the claims environment is changing.

Since the firm obtains its capital and receives premiums at the beginning of a period, but does not pay claims until the end of the period, it must choose how to invest its funds during the period. We assume the investment environment of the capital asset pricing model (CAPM). Within that context, each investor chooses to divide his portfolio among a risk-free security and various risky securities.[3] But equilibrium within the securities market, which is attained after the price of each security has adjusted to equate supply and demand for that security, is shown to imply that each investor purchases the same combination of the available risky securities. In effect, in equilibrium each investor owns a share of the entire market.[4] Thus, investor differences are reflected only in the division of their portfolio between the "risk-free asset" and the "market asset," where the latter contains all of the risky securities in the proportion to their total market value.[5]

At the beginning of the period, the firm has total investible funds equal to $K + PQ$. Its second decision is to choose α, the proportion of these funds that will be invested in the risky ("market") asset, which earns a random rate of return R_m. The remaining proportion $(1 - \alpha)$ is placed in the risk-free asset, which earns a certain rate of return equal to R_f. At the end of the period, the firm's total realized assets then equal

$$(1 - \alpha)(K + PQ)(1 + R_f) + \alpha(K + PQ)(1 + R_m)$$
$$= [1 + (1 - \alpha)R_f + \alpha R_m](K + PQ), \quad (6)$$

where R_m is a random variable with expectation $E(R_m)$, variance denoted by σ_m, and standard deviation denoted by S_m.

Claims costs are not assumed to be necessarily independent of the return on the market asset.[6] We let $\sigma_{m,C}$ denote the covariance between the market rate of return and the claims cost of each policyholder, while ρ denotes the corresponding correlation coefficient:

$$\sigma_{m,C} = \mathrm{Cov}[R_m, C_i] = \rho S_m S_C \qquad \text{for all } i. \quad (7)$$

In the usual model of the insurance firm, risk relates only to the randomness in claims, which are assumed to be independent across policyholders. Thus, the risk of insolvency becomes negligible as the number of policyholders becomes large, if the level of K per policyholder is held constant. In this model, uncertainty about the firm's investment returns is introduced, which does not become negligible as the firm increases in size. Furthermore, the correlation of claims costs across different policyholders —which can represent uncertainty about the probability distribution that applies to claims—means that the variability of the average claim does not necessarily become insignificant as the number of policyholders increases. Thus, the model encompasses three sources of insolvency risk: random claims variability, uncertainly about the parameters of the claims distribution, and variability of investment returns.

Maximizing the Market Value of an Insurance Firm

Let μ denote the difference between the total assets and total liabilities of the firm at the end of the period:

$$\mu = [1 + (1 - \alpha)R_f + \alpha R_m](K + PQ) - \left(\sum_{i=1}^{Q} C_i\right), \quad (8)$$

where K is the initial capital of the firm, PQ is total premium revenue, R_f and R_m are rates of return on the risk-free and the market assets, and C_i is claims costs for the ith policyholder. Taking limited liability into account, the net cash flow μ^+ is as follows:

$$\begin{aligned} \mu^+ &= 0 \text{ if } \mu < 0 \\ &= \mu \text{ if } \mu \geq 0. \end{aligned} \quad (9)$$

Under the CAPM, the market value of the ownership rights to μ^+, valued at the beginning of the period, is

$$V_+ = \left(\frac{1}{1 + R_f}\right)[E(\mu^+) - \theta \operatorname{Cov}(\mu^+, U)], \tag{10}$$

where U represents total cash flow in the entire securities market and θ is a parameter determined by general equilibrium in the market. The term involving the covariance of μ^+ and U represents the market-determined "penalty" for nondiversifiable risk.[7]

The ultimate problem is to analyze which combinations of K, P, Q, and α represent the firm's optimal choices with regard to market value, with limited liability taken into account. However, in the context of limited liability, the question of how to specify the demand relationship arises. Although the assumption of perfect knowledge by applicants is not an interesting context for examining solvency regulation, it would be useful to understand optimal firm behavior in that context.

Since Q, P, K, and α all affect the likelihood of insolvency, we can (conceptually) write demand as a function of these variables. Even beyond the complexity introduced by having Q appear on the right-hand side of the demand relationship $Q = D(Q,P,K,\alpha)$, this specification is too general to yield interesting results. Letting z represent the probability of insolvency, can we simply write $Q = D(P,z)$?

In fact, specifying demand as a function of only P and z neglects the possibility of partial claims recoveries. The condition $\mu < 0$ implies that claims cannot all be fully paid, but only in the extreme case in which total assets are zero at the end of the period will claims recoveries be zero. Thus, the theoretically justified perfect-knowledge assumption is that demand depends on the complete distributions of both final assets and claims. In addition to its complexity, such a specification severely stretches the perfect-knowledge assumption.[8]

Since policyholders are neither perfectly informed nor totally ignorant, the more relevant assumption would be that buyers are partially informed; but to analyze market equilibrium and the effects of regulation with such an assumption is extremely difficult.[9]

An alternative is to assume that applicants use reasonably simple rules, possibly involving proxies for financial solidity. But in that context, a new issue arises: In the context of buyers who use simple rules or proxies to estimate financial condition, firm value cannot be validly analyzed within a one-period model. The basis of firm value becomes fundamentally different. When applicants are assumed able to monitor all relevant parameters and infer their implications for claims outcomes, a firm has no opportunity to build intangible capital. Any firm can enter the market, set P, Q, K, and α, and sell to applicants who completely understand the significance of these choices. For the perfectly informed applicant, a firm's history—including whether it filed for bankruptcy in the previous period—is irrelevant.

Within any context other than perfect knowledge, demand will generally not be a function of only the current values of the parameters. In fact, it is quite reasonable for partially informed applicants, who realize that their inferences about financial condition are unreliable, to view proxies such as the age and size of the firm, and even its advertising budget,[10] as indicators of financial solidity of the firm.[11] But if two firms (in particular, an "old" firm and a "new" firm) that choose the same parameters nevertheless face different demand curves, a source of "goodwill" or "intangible capital" is created. Since a firm can be presumed to lose this intangible capital in the event of insolvency, a one-period model does not capture a crucial aspect of the problem. Since continuing access to the firm's demand curve creates this value, a multiperiod model is required.

We analyze a multiperiod model below, but we are forced to simply set α equal to zero. To gain some insight into the more general problem, we analyze the case of unlimited liability in the remainder of this section. Since policyholders are thus assured of total recovery of all claims, we may assume that demand is a function of only the premium rate, $Q = D(P)$.

The Case of Unlimited Liability

Under the assumptions that the firm's owners are subject to unlimited liability and that they have assets adequate to cover all possible claims, the net cash flow to owners at the end of the period is equal to μ, defined in equation (8), without regard to its sign. Demand is then a function only of the premium rate, $Q = D(P)$. Within this context, we can analyze the choices of Q and α that maximize market value. According to the CAPM, the market value of the ownership of μ is

$$V = \left(\frac{1}{1 + R_f}\right)[E(\mu) - \theta\,\mathrm{Cov}(\mu, U)]. \tag{11}$$

Total market cash flow U includes the additional cash flow generated by the insurance firm under consideration. However, the cash flows represented by the financial investment activity of the firm should obviously not be doubly counted when defining U. Thus, we decompose μ into two components:

$$\mu = \mu^* + \mu^{**}, \tag{12}$$

where

$$\mu^* = (1 + R_f)(K + PQ) - \left(\sum_{i=1}^{Q} C_i\right), \tag{13}$$

$$\mu^{**} = \alpha(R_m - R_f)(K + PQ). \tag{14}$$

Then μ^* can be considered as the additional cash flow generated by the firm,[12] while μ^{**} represents the net cash flow to the firm as a result of its investment in market securities. In other words, if we let M represent all other flows of funds in the securities market at the end of the period, we can write

$$U = M + \mu^*, \tag{15}$$

where μ^{**} is already included in M.

Thus, we can write

$$V = \left(\frac{1}{1 + R_f}\right)\{E(\mu) - \theta \operatorname{Cov}[\mu, U]\} \tag{16}$$

$$= \left(\frac{1}{1 + R_f}\right)\{E(\mu^* + \mu^{**}) - \theta \operatorname{Cov}[\mu^* + \mu^{**}, M + \mu^*]\}, \tag{17}$$

where M is independent of Q and α. Since $E(\cdot)$ and $\mathrm{COV}[\cdot]$ are linear, we can write

$$V = V^* + V^{**}, \tag{18}$$

where

$$V^* = \left(\frac{1}{1 + R_f}\right)\{E(\mu^*) - \theta \operatorname{Cov}[\mu^*, M + \mu^*]\}, \tag{19}$$

$$V^{**} = \left(\frac{1}{1 + R_f}\right)\{E(\mu^{**}) - \theta \operatorname{Cov}[\mu^{**}, M + \mu^*]\}. \tag{20}$$

The total value of the firm is thus the sum of V^* (which represents the value placed by the market on ownership right to the additional funds generated by the insurance firm) and V^{**} (the additional value generated by the investment activity of the firm).

For all Q and all α,

$$V^{**} = 0 \tag{21}$$

(this is proved in appendix A). Since V^* is independent of α, it follows that the market value of the insurance firm is independent of α, the proportion of investible funds placed in the market asset. The reason is that as the firm increases α, the increase in expected earnings is exactly counterbalanced by the penalty imposed by the market on nondiversifiable risk. Since an increase in expected return can be obtained only by increasing the covariance of μ with the market, the higher return on the investment portfolio is exactly offset by the premium demanded by equity owners for bearing additional risk.

Since V^{**} is identically zero and V^* is not a function of α, the value of the firm is maximized when V^* is maximized with respect to Q. Thus, we can write the market value of the insurance firm as

$$V = V^* = \left(\frac{1}{1 + R_{\mathrm{f}}}\right)[E(\mu^*) - \theta \operatorname{Cov}(\mu^*, M + \mu^*)]. \tag{22}$$

In appendix A, we evaluate $E(\mu^*)$ and $\operatorname{Cov}(\mu^*, M+\mu^*)$ and show that

$$V = \left(\frac{1}{1 + R_{\mathrm{f}}}\right)\{(1 + R_{\mathrm{f}})(K + PQ) - QE_C - \theta[-QV_M\sigma_{\mathrm{m},C} + (1 + (Q - 1)\gamma)Q\sigma_C]\}, \tag{23}$$

where V_M represents the total market value of the ownership rights to M, all other end-of-period flows of funds in the market. Taking partials with respect to Q, the first-order condition for maximization of V is

$$(1 + R_{\mathrm{f}})\left(\frac{\partial PQ}{\partial Q}\right) = E_C + \theta[-V_M\sigma_{\mathrm{m},C} + (1 + (2Q - 1)\gamma)\sigma_C]. \tag{24}$$

We can more easily interpret this equation if we translate into different parameters. The general equilibrium within the securities market which is implied by the CAPM can be characterized in two ways. The first, already noted, is a condition relating the value of an asset to the expected value of its gross return μ and the covariance of the gross return with the total flow of funds in the market, U:

$$V = \left(\frac{1}{1 + R_{\mathrm{f}}}\right)[E(\mu) - \theta \operatorname{Cov}(\mu, U)]. \tag{25}$$

Note that the market-determined parameter θ varies with the scale of the total market. An equivalent characterization of the asset market equilibrium can be written in terms of the rate of return on the jth asset, R_j, and the overall rate of return on the entire market, R_U. Equilibrium rates of return satisfy the equation

$$E(R_j) = R_f + \lambda \operatorname{Cov}[R_j, R_U]. \tag{26}$$

The parameter λ, sometimes termed the market risk premium, is equal to

$$\lambda = [E(R_U) - R_f]/\sigma_U, \tag{27}$$

where σ_U is the variance of R_U.[13] The relationship between θ and λ is

$$\lambda = \theta V_U, \tag{28}$$

where V_U refers to the total value of all market assets. Hence, we can rewrite the equilibrium condition as

$$(1 + R_f)\left(\frac{\partial PQ}{\partial Q}\right) = E_C - \lambda\sigma_{m,C}\left(\frac{V_m}{V_U}\right) + \lambda\left(\frac{[1 + (2Q - 1)\gamma]\sigma_C}{V_U}\right). \tag{29}$$

The third term on the right-hand side reflects the increase in total market variance due to the firm in question, and will be essentially zero for all but very large firms. Similarly, we can take $V_m/V_U \approx 1$, except where the firm represents a significant proportion of the total market.

Thus, we can summarize the conditions for maximization of the value of the insurance firm, except where the variability of the firm's returns represents a significant share of the total variance in the entire securities market, as follows:

$$\underset{\alpha, Q}{\text{maximize}}\ V \Rightarrow \left\{\begin{array}{l} \alpha \text{ irrelevant} \\ (1 + R_f)\left(\dfrac{\partial PQ}{\partial Q}\right) = E_C - \lambda\sigma_{m,C} \end{array}\right\}, \tag{30}$$

where V is the market value of the insurance firm, Q is the number of policyholders it accepts at the premium rate P, α is the proportion of investible funds placed in the (risky) market asset, R_f is the rate of return on the risk-free asset, E_C is the expected claims costs per policyholder, λ is the market-determined risk premium, and $\sigma_{m,C}$ is the covariance between the market rate of return and claims costs.

Since $\partial PQ/\partial Q$ is the marginal revenue from adding a policyholder, equation (30) indicates simply that marginal revenue is equal to the discounted value of expected claims cost plus a risk premium which reflects the nondiversifiable component of claims costs. If we assume a competitive insurance market, we can set marginal revenue equal to the premium rate and interpret (30) as determining the competitive premium rate:

$$P = \frac{E_C - \lambda\sigma_{m,C}}{1 + R_f}. \tag{31}$$

Let us consider first the case where claims costs and the market are uncorrelated:

$$\rho = 0 \Rightarrow P = E_C/(1 + R_f). \tag{32}$$

Note that the competitive premium rate is equal to expected cost, per policyholder, discounted at the risk-free rate of return. The fact that the insurance firm can earn an expected rate of return higher than R_f by investing part of its portfolio in the market portfolio does not affect the competitive premium rate. The reason is that the higher return on the investment portfolio must be passed on to equity owners as a return for risk bearing. Exactly the same logic explains why the value of the firm is independent of α.

This equation sheds some interesting light on the long-standing debate over whether investment income should be included in formula for setting premium rates. In a recent hearing on rate setting in Massachusetts, the insurance commissioner argued that investment income should be included in the ratemaking formula. He argued that using the return on "the safest available asset, probably U.S. Treasury Securities, is the minimum standard for investment results in any year . . . . The investment results (of a company that invested only in risk-free bonds) provide a minimum standard for the investment results of real companies. They should on average over time be able to secure better investment returns than the hypothetical company."[14] Equation (31) indicates that even if higher investment returns are earned, they must be passed on to equity owners as a return for risk bearing. Therefore, a premium level set on the assumption of earning the minimum, risk-free rate of return should not be viewed as an upper bound on the level necessary to ensure a competitive rate of return to capital. On the other hand, the insurance firm also does not get a "premium" for bearing the variance related to claims fluctuations. With $\rho = 0$, claims variability is diversified away as the shares of the firm become a trivial fraction of each investor's portfolio.

With $\rho > 0$, the competitive premium rate is reduced below $E_C/(1 + R_f)$, because in that case underwriting income represents an asset with a return that is negatively correlated with the rest of the market. With $\rho > 0$, the insurance business is bad when the market is good, and vice versa; thus the market puts a premium on holding insurance stocks which drives P below E_C. With $\rho < 0$, the reverse applies and the competitive premium rate is above $E_C/(1 + R_f)$. But the effect of ρ on the competitive premium rate has nothing to do with the opportunities for internal portfolio diversification. It is relevant because of the external portfolio diversification opportunities of the firm's owners.

Multiperiod Model with Limited Liability

In the last section, we addressed the problem of how an insurance firm should choose α and Q so as to maximize its own market value. Since the firm's owners were implicitly assumed to face unlimited liability, the amount of invested capital was irrelevant; claimants were paid regardless of the level of K. With limited liability, K becomes a crucial determinant of the probability of insolvency, given the choices the firm makes with regard to other variables. K is therefore a crucial determinant of potential losses to both policyholders and shareholders. In this section we examine the firm's choice of K in the context of limited liability.

Consider first the situation in which applicants are aware of each firm's precise financial prospects. Then applicants will choose the level of safety they prefer, taking the cost of greater safety into account. Where policyholders are assumed to be capable of judging at zero cost the probability of insolvency and its consequences, there are thus no apparent benefits to regulation. Furthermore, an analysis of regulation for that context, whether the regulations are justified or not, would strike most as too unrealistic to be interesting.

In the opening section we discussed the difficulties of relaxing the perfect knowledge assumption. Rather than attempting to propose an "imperfect link" between the firm's choice of those parameters that affect its financial condition and the demand for its policies, we consider the case where demand for the firm's policies is unaffected by the firm's choice of K. This is one method of specifying an insurance-market environment in which applicants have great difficulty evaluating the financial condition and the nature of the risks accepted by insurance firms. It is also relevant to the case in which the policyholder is indifferent to financial condition, either because there is a guaranty fund to compensate the unpaid claimants of insolvent firms or because be purchased (only to qualify as a financially responsible person) insurance covering the liability claims of third parties.

Suppose we apply the assumption that demand for the firm's policies is independent of K in the one-period context. Since neither marginal revenue nor marginal cost is affected by K, the optimal choice for Q will be independent of K.[15] Within a one-period context with the demand curve given, the only function of K is to cover a greater proportion of claims if total claims turn out to exceed total end-of-period assets. Thus, within a one-period model, with limited liability and demand independent of K, the firm will choose to set K as low as possible. If we impose a constraint that K must be positive, we have a "one-cent insurance firm."

But in a multiperiod world, the insolvent firm loses the possibility of earning profits in future periods. The market value of the firm will generally be greater than its liquidation value at the beginning of the period, and, if the firm does not go bankrupt, it will be greater than residual funds at the end of the period. Owing to "goodwill" accumulated through previous service, to simple habit on the part of policyholders, to irrational belief in the solidity of the firm, or to any other phenomenon, the firm's owners possess a stock of intangible capital which is lost in the event of insolvency. We are able to capture this phenomenon by assuming that demand is independent of the firm's choice of K; the firm's intangible capital relates in our context to "access" to the assumed demand curve.[16]

By providing a large amount of paid-in capital K, the firm can protect its access to the market, and specifically its ability to exploit the demand function $Q(P)$. But the larger K is, the less advantage is limited liability if claims turn out to be disastrously high. Thus, in the context of a multiperiod model, the question becomes: Which of two forces is stronger—limited liability, which tends to reduce the optimal level of K to zero, or the desire to protect intangible capital, which is lost when insolvency occurs?

At the beginning of the period, the firm has paid-in capital K and access to the demand function $Q(P)$. Let V represent the market value of the firm at the beginning of the period. We assume that paid-in capital K is fully liquid at full value; thus, the liquidation value of the firm is K and it must be that $V \geq K$. The difference $V - K$ will be termed the intangible capital of the firm and denoted W. Thus,

$$V = K + W, \tag{33}$$

where V is the total value of the firm and W is its intangible capital.

At the end of the period, the firm has net assets of

$$\mu = [1 + \alpha R_m + (1 - \alpha)R_f](K + PQ) - \sum_{i=1}^{Q} C_i. \tag{34}$$

If $\mu \geq 0$, then the ongoing firm retains its intangible capital W but its tangible capital is now μ rather than K.

We assume that the firm is in a stationary environment. Thus, if $\mu \geq 0$, the firm continues to exist and is assumed to face the same demand curve and claims environment at the beginning of the next period. Furthermore, its owners are assumed to have the same investment environment.[17] If $\mu < 0$, the firm is insolvent, defaults on some proportion of its outstanding claims, and loses access to the market. Let us then consider alternative stationary policies, which can be defined thus: If $\mu > K$ at the end of the period, the owners withdraw profits of $\mu - K$; if $0 \leq \mu \leq K$, they provide additional capital of $K - \mu$.

After returning paid-in capital to the level K, the firm's choice environment is identical to that which existed at the beginning of the previous period. Thus, it is obvious that its optimal choices of Q and α will be unchanged from the previous period. Let $Q(K)$ and $\alpha(K)$ refer to the optimal choices for Q and α, given that the firm intends to always return paid-in capital to the level K as long as it remains solvent.

We can thus treat the firm as choosing from among alternative stationary policies $\{K, Q(K), \alpha(K)\}$.[18] Obviously the market value of the firm will depend on the stationary policy chosen, and we can write $V(K)$ to represent the market value of a firm that adheres to such a stationary policy, since K completely defines a stationary policy. With K allowed to vary, clearly $V(K)$ is not the variable that the firm's owners wish to maximize. In fact, the firm's owners will want to continue adding capital as long as the firm's market value increases by more than the addition. That is, the owner's wealth will be maximized if the firm chooses the policy that maximizes $V(K) - K$. If we continue to define "intangible capital" as the difference between the market value of the firm and its liquidation value, so that

$$V(K) = W(K) + K, \tag{35}$$

then the optimal policy is to choose the $\{K, Q(K), \alpha(K)\}$ combination that maximizes the intangible capital $W(K)$ of the firm.

We therefore analyze the nature of $W(K)$, which represents the intangible capital of a firm constrained to follow a stationary policy in which invested capital is always reset equal to K as long as $\mu \geq 0$. Let Y be the market value of the firm at the end of the period, after claims are paid but before capital is restored equal to K. If $\mu < 0$, the firm is insolvent and valueless. If $\mu \geq 0$, then the ongoing firm retains its intangible capital $W(K)$ but its tangible capital is μ rather than K. Thus,

$$
\begin{aligned}
Y &= 0 && \text{if } \mu < 0 \\
&= W(K) + \mu && \text{if } \mu \geq 0.
\end{aligned} \tag{36}
$$

Applying the capital asset pricing model,[19] we can evaluate $V(K)$:

$$V(K) = \left(\frac{1}{1 + R_{\mathrm{f}}}\right)[E(Y) - \theta \operatorname{Cov}(Y, U)]. \tag{37}$$

The covariance term in (37) introduces great complexity, since Y is more likely to be zero if market returns are low. We have not been able to analyze the general case. In order to impose the restriction

$$\operatorname{Cov}(Y, U) = 0 \tag{38}$$

we must assume that claims and market returns are uncorrelated,

$$\rho = 0, \tag{39}$$

and that the firm does not invest in the risky "market security,"

$$\alpha = 0. \tag{40}$$

Either $\rho \neq 0$ or $\alpha \neq 0$ implies that $\operatorname{Cov}(Y, U) \neq 0$.

With the assumptions $\rho = 0$ and $\alpha = 0$, we can write[20]

$$V(K) = W(K) + K = \left(\frac{1}{1 + R_{\mathrm{f}}}\right)E(Y). \tag{41}$$

Let $\delta(K)$ represent the critical level for $\bar{C}$ (average realized claims), at which the firm becomes insolvent:

$$\delta(K) = (K + PQ)(1 + R_{\mathrm{f}})/Q, \tag{42}$$

so that $\mu < 0$ when $\bar{C} > \delta$. Then $Y = 0$ when $\bar{C} > \delta$, and we can write

$$E(Y) = E[\mu + W(K) | \bar{C} \leq \delta(K)]. \tag{43}$$

Let $f(\cdot)$ represent the probability density for $\bar{C}$:

$$\bar{C} \sim f(\bar{C}). \tag{44}$$

We can then write

$$E(Y) = \int_{-\infty}^{\delta} [W(K) + (K + PQ)(1 + R_{\mathrm{f}}) - Q\bar{C}] f(\bar{C}) d\bar{C}. \tag{45}$$

Let F_δ be the probability that the firm becomes insolvent during any single period:

$$F_\delta = \int_{\delta(K)}^{+\infty} f(\bar{C})d\bar{C}, \tag{46}$$

and let E_δ be expected claims costs, taking limited liability into account. That is, E_δ equals expected claims costs, conditional on $\mu \geq 0$:

$$E_\delta = \int_{-\infty}^{\delta(K)} \bar{C}f(\bar{C})d\bar{C}. \tag{47}$$

Combining these definitions with (45), we can write

$$V(K) = \frac{[W(K) + (K + PQ)(1 + R_f)](1 - F_\delta) - QE_\delta}{1 + R_f}. \tag{48}$$

Combining (41) and (48), we can solve for $W(K)$:

$$W(K) = \frac{PQ(1 + R_f)(1 - F_\delta) - K(1 + R_f)F_\delta - QE_\delta}{R_f + F_\delta}. \tag{49}$$

We now analyze the behavior of $W(K)$ as K varies.[21] First, W(K) does not grow beyond bound. Since F_δ, the probability of insolvency, becomes zero as K grows while E_δ approaches E_C, it follows that[22]

$$W(\infty) = \frac{PQ(1 + R_f) - QE_C}{R_f} = \frac{(P(1 + R_f) - E_C)Q}{R_f}. \tag{50}$$

Clearly, $W(\infty)$ represents the profits of $P(1 + R_f) - E_C$ per policyholder which are earned with certainty each period when $K = \infty$.

We show in appendix B that if $W(K)$ has a positive slope for $K = K_1$, then it must have a positive slope for all K such that $K > K_1$, as long as the distribution of average claims falls within a wide class of distributions which include the range of plausible specifications.[23] It thus follows that $W(K)$ must follow one of the following four patterns:

Case 1
$W(K)$ decreases montonically toward $W(\infty)$.

Case 2
$W(K)$ decreases initially, then increases toward $W(\infty)$, with $W(0) > W(\infty)$.

Case 3
$W(K)$ decreases initially, then increases toward $W(\infty)$, with $W(0) < W(\infty)$.

Case 4
$W(K)$ increases monotonically toward $W(\infty)$.

We illustrate these four cases in figure 3.1.

Consider the firm that is free to set K at any level. This exhaustive list of cases indicates that an internal solution with $0 < K < \infty$ never exists. There are only two possibilities: When $W(0) > W(\infty)$, the firm will wish to put no capital into the firm; when $W(\infty) > W(0)$, intangible capital $W(K)$ can always be increased by adding more capital.[24]

This result is not surprising; it reflects the competing desires to take advantage of limited liability and to protect the intangible capital of the firm. If the value of access to the market is too small, then the lure of limited liability dominates. The optimal policy in that case is to continue in business only as long as total claims are "favorable" (that is, $\bar{C} < P(1 + R_f)$) and to "plan" to become insolvent in the first period in which total claims are "unfavorable." Although the firm may have intangible capital as a going concern, represented by $W(\infty)$, the value of guarding it is less than the potential value of taking advantage of limited liability.

When $W(\infty) > W(0)$, the optimal policy is to set $K = \infty$. If the value of access to the demand curve $Q(P)$ is sufficiently great, then the addition of capital, by increasing the expected longevity of the firm, increases the market value of the firm by more than the value of the added capital. But as the firm's owners supply more capital, the probability of insolvency continues to fall and becomes negligible. As capital is added, the firm looks more and more like an infinitely safe financial intermediary. The "cost" of adding capital continues to fall. Within the assumptions of the model, the value of the firm continues to increase by more than the value of the added capital, and thus there is no finite solution for K.[25]

The absurdity of the "solution" $K = \infty$ reflects the limited realism of the model. Various real-world considerations imply that the supply price of capital to be added to firm reserves would eventually rise. The risk of embezzlement, other fiducial risks, and the desire of investors to diversify across different investment managers, would all act in that direction. But the principle remains: The insurance firm with an incentive to add capital and to drive the risk of insolvency to a low level can then continue to add to reserves at a very low cost and thereby drive that risk to a negligible value. Thus, when we refer to the "solution" $K = \infty$, we refer to a situation in which capital has been added until these problems of agency responsibility dominate. If the "fiducial responsibility" risks are not great, as would be expected for trusted managers, we can conclude that a "realistic"

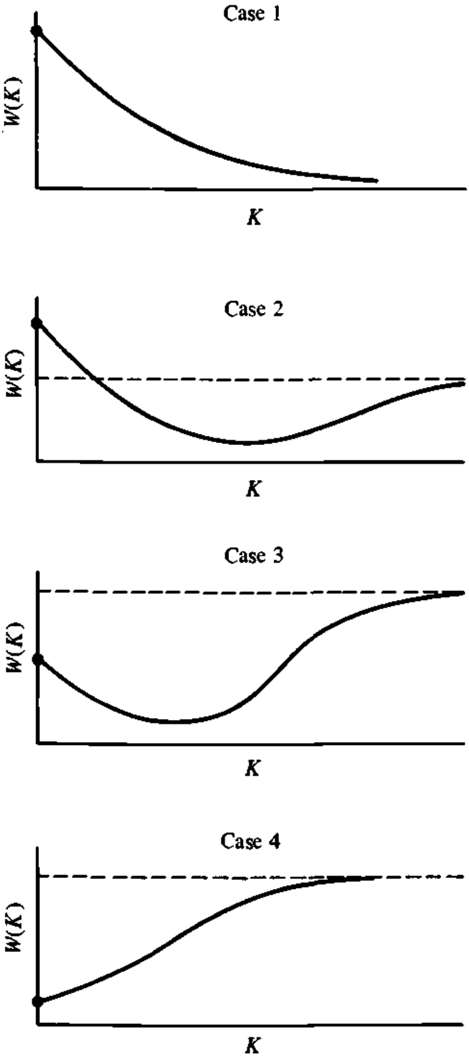


Figure 3.1 Four possible shapes for $W(K)$.

Table 3.1 Illustrative values of ξ and the corresponding threshold for R_f.

ξ	$R_f = \psi(\xi)$
0.000	0.000
0.025	0.032
0.039	0.050
0.050	0.064
0.075	0.097
0.077	0.100
0.100	0.131
0.114	0.150
0.149	0.200
0.150	0.201
0.200	0.274
0.250	0.350
0.300	0.430
0.400	0.598
0.500	0.780

solution for this case is characterized by a value for K large enough that the probability of insolvency is negligible.

The ratio of $W(\infty)$ to $W(0)$, thus, is a crucial determinant of firm behavior with respect to K. We have been unable to obtain any perfectly general results, but we have derived an interesting result for the case in which average claims are assumed to be normally distributed.[26] Thus, we now assume that $\bar{C}$ is normally distributed with mean E_C and standard deviation $\bar{S}$:[27]

$$\bar{C} \sim N(E_C, \bar{S}_C). \tag{51}$$

Let us express the premium rate, inflated at the risk-free rate of return, in terms of E_C and $\bar{S}_C$:

$$P(1 + R_f) = E_C + \xi \cdot \bar{S}_C. \tag{52}$$

Thus, if $\xi = 0.10$, the inflated premium rate is 10 percent of one standard deviation above the mean of the distribution of average claims.

We show in appendix B that the relationship between $W(0)$ and $W(\infty)$ depends only on ξ and R_f; specifically,

$$W(0) > W(\infty) \qquad \text{for } R_f > \psi(\xi),$$

$$W(\infty) > W(0) \qquad \text{for } R_f < \psi(\xi),$$

Table 3.2 Critical levels for $[P(1 + R_f) - E_C]/S_C$.

	Q		
R	100	1,000	10,000
With $\delta = 0$:			
0.05	0.0039	0.0012	0.0004
0.10	0.0077	0.0024	0.0008
0.15	0.0114	0.0036	0.0011
0.20	0.0149	0.0047	0.0015
With $\delta = 0.10$:			
0.05	0.0129	0.0124	0.0124
0.10	0.0254	0.0245	0.0244
0.15	0.0375	0.0361	0.0359
0.20	0.0493	0.0474	0.0472

where ψ is a function which relates to the normal distribution.[28] Over the interesting range, $\psi(\xi)$ is between 25 and 50 percent greater than ξ. Illustrative values of $\psi(\xi)$ are shown in table 3.1. For $\xi = 0.10$, $\psi(\xi) = 0.131$. Thus, if the inflated premium $P(1 + R_f)$ is 10 percent of $\bar{S}_C$ above E_C, then the critical value for R_f is 0.131. With $R_f < 0.131$, the firm will want to place an infinite amount of capital in the firm, but if $R_f > 0.131$ the firm will want to withdraw all of its capital. Conversely, with $R_f = 0.20$, the critical value for is ξ is 0.149.

Since ξ refers to the margin between the inflated premium rate and a standard deviation of the average claims distribution, it appears clear from table 3.1 that only a narrow margin is necessary for the firm to protect its market access, for firms of even moderate size. In table 3.2, for instance, we translate the implied margins into units in S_C rather than in $\bar{S}_C$. Thus we compute the critical level for $[P(1 + R_f) - E_C]/S_C$ above which the firm will want to set $K = \infty$, where S_C is a standard deviation for the claims distribution of each policyholder, rather than for the average claims distribution.

With $\delta = 0$, the critical margin between $P(1 + R_f)$ and E_C is much less than 1 percent of a standard deviation, unless the firm is very small or the risk-free rate of return is very high.[29] Even with 100 policyholders and $R_f = 0.20$, the critical level is only 1.5 percent of a standard deviation.

If we allow the claims of different policyholders to be correlated, however, the critical margin increases significantly and the effect of firm size is virtually eliminated. As noted above, the case of $\gamma \neq 0$ can be interpreted as reflecting uncertainty about the location of the claims distribution. Table 3.2 also shows the margins with $\gamma = 0.10$.[30] The margins then range from 1 percent to 5 percent and are affected only slightly as Q ranges from 100 to 10,000 policyholders.

With the product $Q\delta$ is even moderately large (>200), we can use the following excellent approximation for the critical level ξ:

$$[P(1 + R_f) - E_C]/S_C \approx \xi\sqrt{\delta}. \tag{53}$$

Thus, even when δ approaches 1 the critical margin does not exceed ξ. When $R_f = 0.20$, the critical margin thus does not exceed 0.149, regardless of the level of δ and the source of the variation in the average claims distribution.

Regulating Minimum Capital

We now consider the effect of regulating the level of paid-in capital. Suppose the firm is constrained by a requirement that $K \geq \bar{K}$. The effect will depend on the behavior of $W(K)$. In figure 3.2 we add a constraint on K to the four cases we have identified. For case 3 and case 4, in which $W(\infty) > W(0)$ and the unconstrained firm will choose $K = \infty$, the constraint is clearly not binding and the regulation will have no effect.

With case 1, in which $W(K)$ is monotonically decreasing, the regulated firm will clearly choose $K = \bar{K}$, the constraint will be binding, and the firm will choose to keep only as much capital as required by regulation.

Case 2 provides an interesting possibility. With $K = \bar{K}_1$, such that $W(\bar{K}_1) > W(\infty)$, the firm will choose to set $K = \bar{K}_1$, and the constraint will be binding, just as with case 1. But suppose that $K = \bar{K}_2$, for which $W(\bar{K}_2) < W(\infty)$. In this case the optimal policy for the firm which is required to have $K \geq \bar{K}_2$ is $K = \infty$. Although the unconstrained firm would choose to set $K = 0$, the constrained firm would choose $K = \infty$.

It is thus of interest to determine when the different cases occur. Although we have not found a simple characterization of the parameter combinations that produce the four cases, calculations of $W(K)$ for a wide range of parameter values show a consistent and plausible pattern. Cases 1, 2, 3, and 4 occur in that order as ξ is increased from $\xi = 0$, with the other parameters held constant. Thus, if the claims distribution and the interest rate R_f are held constant, the nature of $W(K)$ is successively represented by cases 1, 2, 3, and 4 as the inflated premium $P(1 + R_f)$ is increased from expected claims cost E_C to larger values. Because Q is being held constant, these calculations implicitly refer to a shifting demand curve, or, in a competitive market, to an increasing competitive premium rate.

We demonstrate in appendix B that $W(K)/Q$ is completely determined by the parameters E_C, $\bar{S}_C$, R_f, ξ, and s, where

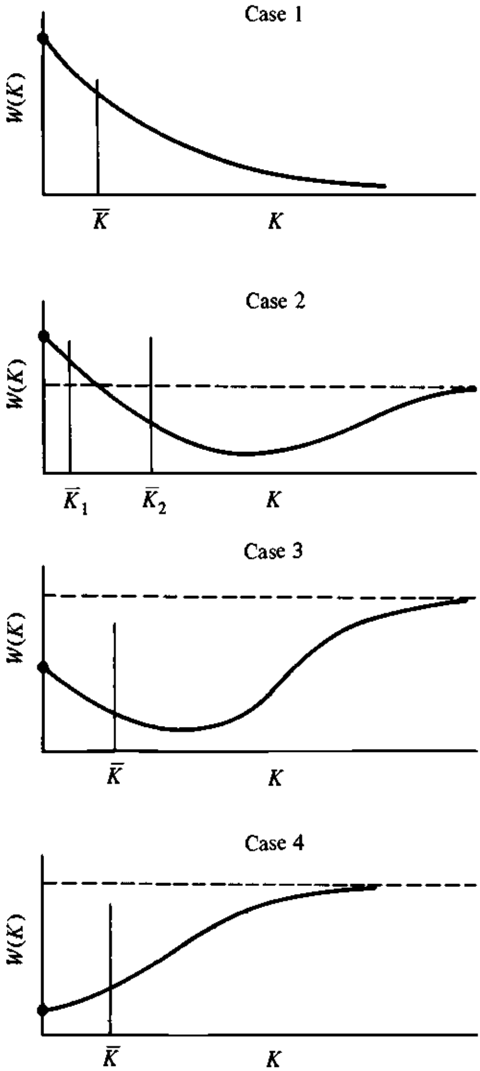


Figure 3.2 Imposing a minimum-capital constraint in the four cases.

$s = K/PQ.$

Figure 3.3 shows $W(K)/Q$ for values of s ranging from 0 to 0.20, with the other parameters fixed at illustrative values. For this set of parameters, the four cases occur as follows:

Case 1	$\xi = 0$
Case 2	$\xi = 0.025, 0.050, 0.075$
Case 3	$\xi = 0.100, 0.150, 0.200$
Case 4	$\xi = 0.50$ (not shown on figure).

Over the range of parameters for which $W(K)/Q$ was calculated,[31] the same sequence for the four cases was followed consistently;[32] however, we have no proof that it will hold for every parameter combination.

Insurer Behavior: A Partial Integration

In the preceding section we analyzed the firm's optimal policy with respect to K, treating its choices of the premium (P) and the number of policyholders (Q) as given. We found that the firm's behavior with respect to invested capital depends on the relationship among the premium rate, the average claims distribution, and the risk-free rate of return. Suppose the inflated premium rate is expressed in units that refer to the number of standard deviations by which $P(1 + R_f)$ lies above the mean of the claims distribution. That is, let ξ be defined such that

$$P(1 + R_f) = E_C + \xi \cdot \bar{S}_C, \tag{54}$$

where E_C and $\bar{S}_C$ are the mean and the standard deviation of the distribution of average claims costs. We find that if

$$\psi(\xi) < R_f \tag{55}$$

then the unregulated firm will choose to set $K = 0$, but if

$$\psi(\xi) > R_f \tag{56}$$

the firm will choose to set $K = \infty$, where $\psi(\cdot)$ is a function derived from the density function of the normal distribution.[33] Although the "solution" $K = \infty$ is not literally acceptable, this case appears to be validly interpreted as a solution for which the probability of insolvency is reduced to a negligible value and can be treated as essentially zero.

These results refer to the optimal choice for K with P and Q given, with investment opportunities assumed irrelevant,[34] and with the demand

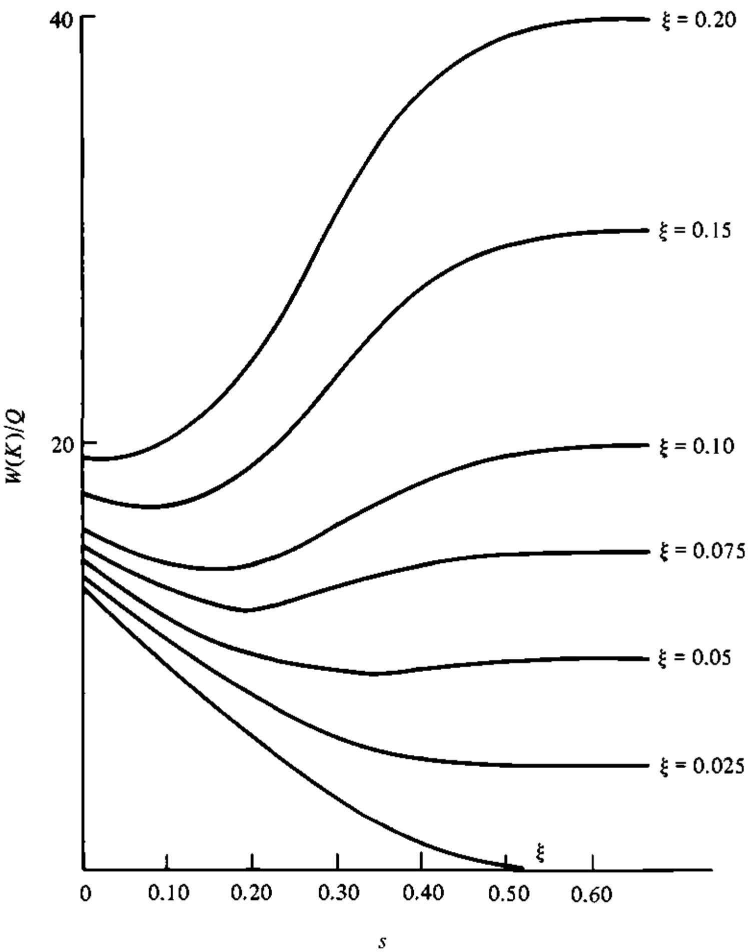


Figure 3.3 $W(K)/Q$ as a function of s for different values of ξ, with $E_C = 100$, $\bar{S}_C = 20$, and $R_f = 0.10$.

relationship $P(Q)$ assumed independent of K. The issue arises: Which combinations represent joint optima, under these conditions, when the firm simultaneously maximizes with regard to P, Q, and K?

Let $W(P,Q,K)$ represent the net, or intangible, value of the firm.[35] The fundamental relationship

$$\underset{P,Q,K}{\text{maximum}}\ W(P,Q,K) = \underset{P,Q}{\text{maximum}}\ \{\underset{K}{\text{maximum}}\ W(P,Q,K)\} \tag{57}$$

holds generally. However, as demonstrated in the preceding section, the problem

$$\underset{K}{\text{maximize}}\ W(P,Q,K) \tag{58}$$

results in a solution of either $K = 0$ or $K = \infty$ for all possible P and Q. Thus, if we obtain solutions, subject to the constraint represented by the demand relationship, for the two problems

$$\underset{P,Q}{\text{maximize}}\ W(P,Q,O) \tag{59}$$

$$\underset{P,Q}{\text{maximize}}\ W(P,Q,\infty), \tag{60}$$

we can, in theory, find all joint equilibria by finding solutions that are mutually consistent. That is, suppose we can find triples P^*, Q^*, K^* such that K^* is optimal, given P^* and Q^*, while P^* and Q^* are optimal, given K^*. Such a combination represents a joint equilibrium and a local optimum.[36] The global joint optimum must be one of these local joint optima.

But the analysis for problem (60) has been done. When the firm sets $K = \infty$, limited liability obviously becomes irrelevant. Thus, the analysis of the first section ("One-Period Model . . ."), restricted to the case of $\alpha = 0$ and $\rho = 0$, applies. The optimal values for P and Q derived in that section are independent of the value assumed for K; this follows from the implicit assumption of unlimited liability. Clearly, if we now assume that K is sufficiently large that limited liability is irrelevant, the same solution results.

Under the restriction $\rho = 0$, and thus $\sigma_{m,C} = 0$, the optimal solution obtained in the first section is simply the premium and policies-written combination for which marginal revenue equals discounted expected claims cost:

$$MR = \left(\frac{1}{1 + R_f}\right)E_C. \tag{61}$$

Let P^* and Q^* represent the values that satisfy (61). Further, let ξ^* denote the corresponding value for ξ:

$$P^*(1 + R_f) = E_C + \xi^* \cdot \bar{S}_C. \tag{62}$$

Then there are two cases to consider. Suppose first that, for this combination,

$$\psi(\xi^*) > R_f. \tag{63}$$

Then it is clear that the combination $[P^*, Q^*, K = \infty]$ represents a local joint optimum. The choices P^* and Q^* imply that $K = \infty$ is optimal, while $K = \infty$ implies that P^* and Q^* are optimal. In this case the probability of insolvency is negligible and policyholders effectively enjoy complete protection.

Now suppose instead that

$$\psi(\xi^*) < R_f, \tag{64}$$

so that, with P^* and Q^* given, the firm would choose $K = 0$. It does not follow that the combination $[P^*, Q^*, K = 0]$ is a local optimum, since the optimality of P^* and Q^* presumes unlimited liability. With $K = 0$, the optimal choices for P and Q obviously must reflect limited liability. Although we have not made a complete analysis of P^{**} and Q^{**}, the optimal choices with $K = 0$, it must be true that

$$\psi(\xi^{**}) < R_f, \tag{65}$$

where ξ^{**} corresponds to P^{**}.[37] Thus, in the case for which $\psi(\xi^*) < R_f$, we know that some P^{**} and Q^{**} exist such that the combination

$$[P^{**}, Q^{**}, K = 0]$$

represents a global optimum for the simultaneous determination of all three variables. Policyholders face a risk of insolvency. There is no bound on the probability of insolvency, at least with the assumptions given. A demand curve and a claims cost distribution could be defined for which the firm would choose a probability of insolvency arbitrarily close to unity.

For the case in which $\psi(\xi^*) > R_f$, if does not necessarily follow that the combination

$$[P^*, Q^*, K = \infty]$$

is a global optimum, even though it is a local optimum. It may be that a combination exists with $K = 0$ which dominates it. In essence, the firm

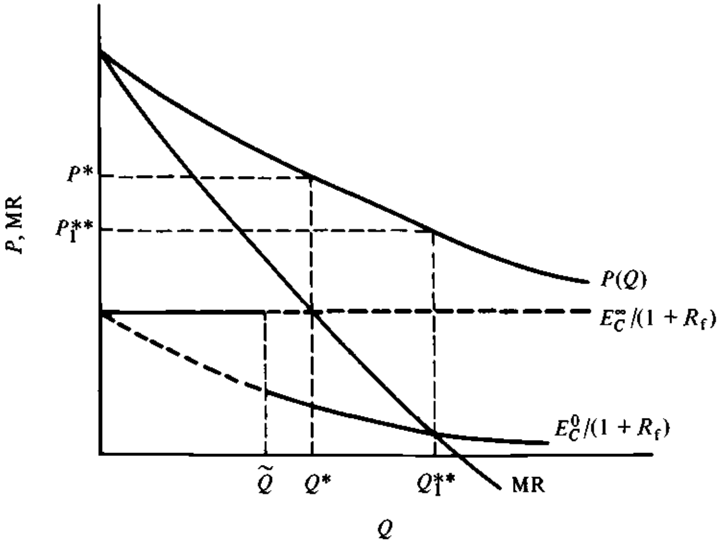


Figure 3.4 Equilibrium with $Q^* > \tilde{Q}$.

has a choice of operating with the marginal-cost curve E_C with assured survival, or of operating with a lower marginal-cost curve by setting $K = 0$. Of course, by choosing the lower marginal-cost curve, it accepts a lower probability of survival each period. Whether the local optimum represented by P^*, with $K = \infty$, or the local optimum represented by P^{**}, with $K = 0$, represents a higher level for intangible capital depends on the precise nature of the demand curve.

The different cases can be clarified diagrammatically. Limited liability effectively reduces marginal expected policyholder costs, with the effect on marginal cost becoming greater as Q increases and the probability of insolvency grows.[38] In figure 3.4 we depict marginal expected policyholder cost with $K = 0$, labeling it E_C^0, while using E_C^∞ to label "full" expected marginal cost,[39] corresponding to $K = \infty$. We let $\tilde{Q}$ correspond to $\tilde{\xi}$ such that

$$\psi(\tilde{\xi}) = R_f, \tag{66}$$

so that the firm will choose to set $K = \infty$ if $Q < \tilde{Q}$, and will choose to set $K = 0$ if $Q > \tilde{Q}$.[40] Thus, the expected-marginal-cost curve E_C^∞ applies for $Q < \tilde{Q}$, while E_C^0 applies for $Q > \tilde{Q}$.

The case in which $Q^* > \tilde{Q}$, and thus $\psi(\xi^*) < R_f$, is illustrated in figure 3.4. It is clear that the solution for the one-period problem corresponds to the point at which marginal revenue equals $E_C^0/(1 + R_f)$. We

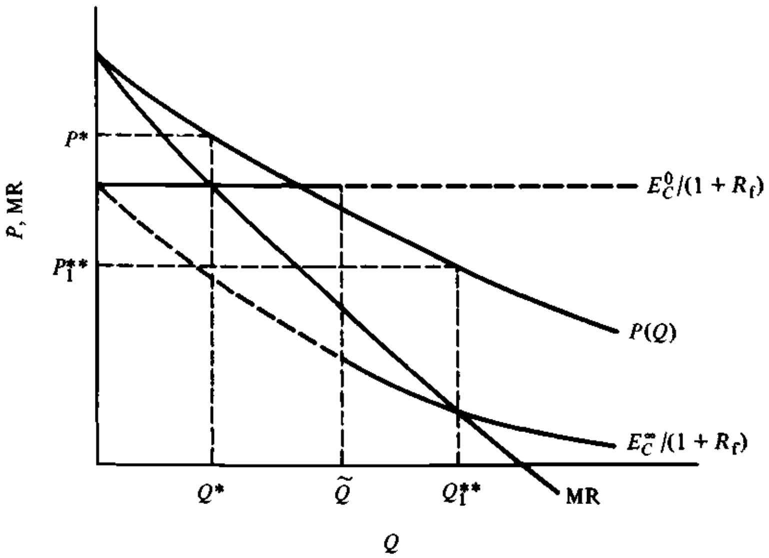


Figure 3.5 Alternative equilibria with $Q^* < \tilde{Q}$.

use Q_1^{**} and P_1^{**} to label the one-period solution. Note that P_1^{**} could be less than E_C^∞; in the case of limited liability, the firm could choose to write policies below "full" expected costs.

But within a multiperiod context, the value of survival becomes relevant. Decreasing P and increasing Q both decrease the probability of survival. If P^{**} denotes the optimal premium for the multiperiod case with $K = 0$, it is clear that P^{**} will be greater than P_1^{**}.[41]

Figure 3.5 illustrates the case for which $Q^* < \tilde{Q}$ and $\psi(\xi^*) > R_f$. The two possible one-period equilibria are indicated, and it is clear that the global optimum will depend on the elasticity of the demand curve at prices below P^*. The relevant comparison, of course, is between $[P^*, Q^*, K = \infty]$ and $[P^{**}, Q^{**}, K = 0]$. If the demand curve is sufficiently elastic, the profits expected to be made by increasing Q, setting $K = 0$, and taking eventual advantage of limited liability dominate the value of assured survival.

A global equilibrium with regard to the simultaneous choice of P, Q, and K can thus be characterized by three possible cases. If $\psi(\xi^*) < R_f$, it follows that the firm will choose to set $K = 0$ if it has complete discretion. It is possible that P may be chosen to be below $E_C(1 + R_f)$. Policyholders face a risk of insolvency which, at least given our assumptions, may be arbitrarily high. If a guaranty fund protects claimholders, then the deficits are shifted to the industry as a whole.

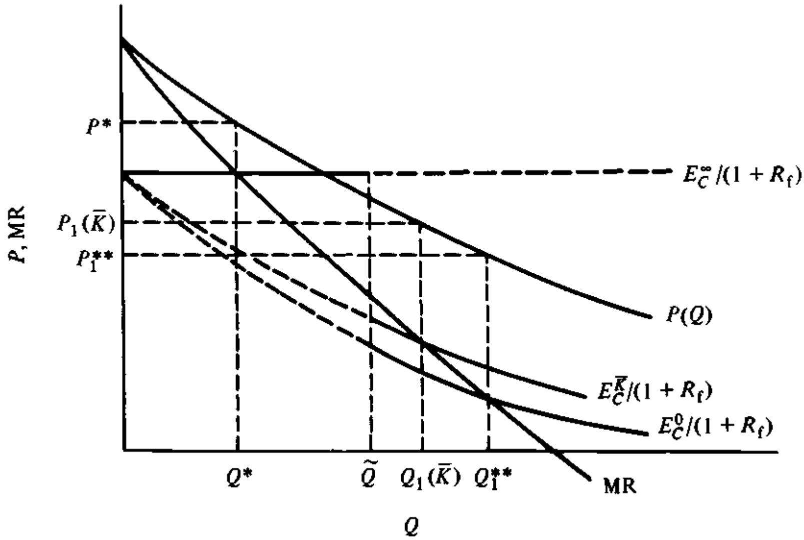


Figure 3.6 Equilibrium with $Q^* < \tilde{Q}$ and $K \geqslant \bar{K}$.

If $\psi(\xi^*) > R_f$, then two cases are possible. It may be that P^* and Q^*, as defined by (61) are optimal along with $K = \infty$. But if the demand curve is sufficiently elastic below P^*, the firm may attain a still higher market value by setting $K = 0$ and adjusting its premium and number of policies written. We can prove that the choice for P will be below P^* and that the policies written will be greater than Q^*.

Regulation and Minimum Required Capital

We can now consider the effect of regulating minimum K when P, Q, and K are simultaneously chosen to maximize the net market value of the firm. Obviously, in the case in which the optimal strategy for the firm (in the unregulated context) is to set $K = \infty$, imposing the requirement $K \geq \bar{K}$ has no effect. Within the context of our assumptions, the costs of regulation are entirely administrative and enforcement costs. Costs of inefficiency that arise as firms adapt to regulations do not occur.

As we have demonstrated, the only alternative case is that in which the unregulated firm would choose a combination $[P^{**}, Q^{**}, K=0]$. We saw above that when the constraint $K \geq \bar{K}$ is introduced, with P and Q held constant, the firm may choose to set $K = \bar{K}$ or may choose to set $K = \infty$. When P and Q are allowed to vary, these same cases are possible.

Imposing a minimum capital requirement rotates the expected-marginal-cost curve toward E_C^∞. Figure 3.6 illustrates, for one value of $\bar{K}$, the case in which the unregulated firm would choose $[P^{**}, Q^{**}, K=0]$ even though $\psi(\xi^*) > R_f$. For the one-period solution, the regulated firm clearly might choose either $[P^*, Q^*, K=\infty]$ or the combination $[\bar{P}, \bar{Q}, K=\bar{K}]$. Obviously, the same possibilities exist for the multiperiod case.

It is also clear that, for $\bar{K}$ large enough, the firm will choose $[P^*, Q^*, K=\infty]$. However, for the case in which $\psi(\xi^*) < R_f$, the firm will never choose to switch over to $K = \infty$.

Thus, we can characterize the effects of regulation in terms of the two basic cases. For $\psi(\xi^*) < R_f$, the unregulated firm will choose to set $K = 0$. The regulated firm, which is required to set $K > \bar{K}$, will choose to set $K = \bar{K}$. Furthermore, it will also respond by setting a higher premium rate P and reducing the number of policies written. Thus, for a firm that would choose to "exploit" limited liability, the imposition of a minimum capital requirement reduces the level of risk accepted, in addition to providing a cushion for those policies written.

For $\psi(\xi^*) > R_f$, there are two possibilities. The unregulated firm may choose to set $K = \infty$. The unregulated firm would voluntarily choose to be "infinitely safe," and regulations are neither necessary nor relevant. The other possibility is that the unregulated firm would choose a combination $[P^{**}, Q^{**}, K=0]$. If a small minimum capital requirement is imposed, the firm will react as above: by setting $K = \bar{K}$, raising the premium rate, and reducing Q. However, for a sufficiently high $\bar{K}$, the firm will make a discontinuous jump to the combination $[P^*, Q^*, K=\infty]$. Thus, minimum capital requirements can generate "infinitely safe" firms in some cases. But "infinitely safe" firms also may exist in the absence of regulation, as a result of the desire to protect access to a profitable market.

Solvency Regulation: Repairing a Deficit

One aspect of the analysis requires clarification. We have assumed, in the analysis of the case of limited liability, that the firm is faced with the following possibilities: If total funds μ at the end of the period exceed total claims, the firm pays all claims, repairs capital to its beginning level, and begins the next period; if total claims exceed total funds, then all remaining funds are used to pay claims and the firm is then declared insolvent.

This set of assumptions is consistent with the power of state insurance commissioners to liquidate any firm whose capital and surplus falls below the required level. But one important possibility is left out: the possibility that the firm's owners will be given the opportunity to "resurrect" the firm by adding capital at the end of the period. Clearly, "resurrection" may or may not be in the interests of the owners, depending on just how large the deficit is.

Owners cannot be forced to repair a deficit, given limited liability. If they are always allowed to repair a deficit, then the optimal strategy with regard to invested capital, within this model, is obvious. With demand totally insensitive to K, the firm will always choose to set $K = \bar{K}$. The only function of "excess" capital in that context is to pay a greater proportion of claims in the event that claims are sufficiently high, in some period when it is not worthwhile to salvage the firm.

This situation presents the regulator with a dilemma. The regulator will always want to allow the owners to salvage the firm, *ex post*, if he ignores the effect on future incentives. But if he always allows owners to salvage when they wish, the incentive to provide the firm with more than the minimum required capital disappears. On the other hand, the owners are more likely to want to salvage the firm if they know they can choose whether to salvage in the future. The firm is worth more if its owners have a right to salvage.

Where other aspects of the cost of capital are ignored, as we have done here, the answer seems clear. Since in our context the cost of capital relates only to the uncertainty about claims variations, and becomes negligible as the probability of insolvency becomes small, the regulator should provide the incentive to set $K = \infty$. But these other aspects of the supply of capital impinge not only on the costs of existing and established firms, but also on the difficulty of entry and on the startup costs of new and young firms. The effects of solvency regulation seem to be clearest with respect to the total supply of firms in the industry, rather than with respect to insolvency rates (Munch and Smallwood 1980). Thus, one should be cautious in drawing policy conclusions from the analysis.

Summary and Conclusions

We first examine optimal firm choices in the context of unlimited liability. We find that α, the proportion of initial funds invested in the risky asset, does not affect the market value of the firm. Under the assumptions of the CAPM, the insurance firm's owners are indifferent with regard to the investment behavior of the insurance firm. Although the firm's expected rate of return will be higher if it invests in the risky asset, the securities market requires a compensating premium to the extent that the additional risk is nondiversifiable.

The optimal choice for P also reflects the extent to which investors can diversify away the variance in the firm's cash flow. The optimal P and Q are defined by

$$(1 + R_f)\text{MR} = E_C - \lambda\sigma_{m.C}. \tag{67}$$

That is, inflated[42] marginal revenue is equal to expected policyholder cost plus a term representing compensation for nondiversifiable risk. For a competitive market, we can interpret equation (67) as defining the competitive premium rate. Thus, investment opportunties are relevant to the level of competitive premiums, but the link involves the external investment opportunities of investors and will operate through the cost of capital to the firm. The fact that the insurance firm can invest its capital and reserve funds in assets which earn an expected return above R_f is irrelevant to the optimal choice of P and Q (for a firm with market power) and to the premium level in a competitive market.

It is important to note that the indifference of owners to the firm's investment behavior is not based on an assumption of investor risk neutrality. Within the CAPM, the community of investors are assumed to be risk-averse, and they can be assumed to be risk-averse to an arbitrarily high degree. The point is that their own portfolio diversification renders the investment behavior of the firm irrelevant.

These conclusions may be considered unrealistic, but it is important to recognize the nature of the phenomenon that must be modeled to make the analysis significantly more realistic. Three assumptions implicit in the CAPM drive these conclusions. The first is that securities trading does not involve transactions costs, in the sense that ownership may be transferred without cost. The second is that all potential investors share common information about the distribution of returns to all firms. The third is that the returns of each firm are given, and implicitly that the behavior of managers is thus also given and unaffected by changes in ownership. Let us call these the problems of *ownership transfer*, *information*, and *agency*, respectively.

Although the costs of ownership transfer clearly affect the small investor, it seems implausible that the market value of firms would be highly sensitive to such costs in a world in which the information and agency problems did not exist. The existence of large traders and financial intermediation would tend to eliminate its effect for all but very small firms. Thus, to improve significantly on the CAPM analysis of market value, particularly for the case of closely held firms we must address the information and agency problems.

The agency problem relates to the classic issues involving incentives and control when management is divorced from ownership. One might presume that managers will want to minimize the possibility of insolvency, but this depends on the nature of their remuneration and on their risk aversion.

With regard to information, managers may well have a better understanding of the relationships among claims costs, returns on securities, and the regulations specific to the forms in which reserves must be held. Thus, the external portfolio management of owners may not be perfectly substitutable for internal portfolio management, and the firm's investment behavior may not be irrelevant to market value.

If the problems of agency and information were explicitly modeled, the analytical results would quite likely be modified somewhat, particularly for the closely held firm. But we still believe them to be first-order solutions that are fundamentally correct.

Analysis of a firm's optimal choices in the context of limited liability presents a fundamental problem: How should demand for the firm's product be specified, given the possibility of insolvency? Since the case of perfect knowledge is uninteresting, the problem becomes one of specifying the response of partially informed buyers. But there is no obvious and convincing method for specifying the relationship between the firm's choices and the judgments that partially informed prospective policyholders make about the solidity of the firm.

Rather than analyze some arbitrary specification of the link between demand and the firm's choice for these variables, we analyze the case in which demand is sensitive to only the premium rate. For the case in which policyholders are protected by a guaranty fund providing claims coverage when the insurer becomes insolvent, this seems the appropriate assumption in any case. Where no guaranty fund exists, it is still useful to understand the nature of firm behavior when demand is completely insensitive.

Even with demand taken as a function of only the premium rate, an analysis of the joint determination of K, α, P, and Q in the context of limited liability is complex, primarily because a one-period model is no longer adequate. When buyers are less than perfectly informed, their behavior and beliefs represent sources of intangible capital that is lost when insolvency occurs. In our model, the firm's intangible capital relates to access to the demand curve $Q(P)$, which the firm is assumed to lose when insolvency occurs.

Under a multiperiod model, market value is determined by the distribution of returns and by the probability of survival each period. In order to keep the analysis tractable, we impose the restrictions $\alpha = 0$ and

$\rho = 0$. Thus, the role of investment behavior is assumed away.[43] For the case of limited liability, we consider only the firm's choices for K, P, and Q.

The optimal choice for K (and thus the probability of survival) is initially analyzed with a given distribution of returns (with P and Q fixed). Market value is determined for alternative stationary policies in which the owners reestablish K at a specific level each period as long as the firm survives. The level of intangible capital, or market-value net of K, is demonstrated to be a U-shaped function that monotonically approaches an asymptote as K grows. This shape reflects two competing forces: a high level for K ensures survival, whereas a low level allows the firm to exploit limited liability in the event of large claims.

The optimal policy is thus found to consist of two extremes: The firm should choose to set $K = 0$ or $K = \infty$, depending on its profitability and on the rate of interest. Specifically, let ξ be the number of standard deviations of the distribution of average claims cost by which the inflated premium rate lies above expected average claims cost:

$$P(1 + R_f) = E_C + \xi \bar{S}_C. \tag{68}$$

The optimal choice for K can then be characterized as follows:

$$\begin{aligned} K &= 0 \Leftrightarrow \psi(\xi) < R_f, \\ K &= \infty \Leftrightarrow \psi(\xi) > R_f, \end{aligned} \tag{69}$$

where $\psi(\cdot)$ is a function that relates to the normal distribution.[44] Though the "solution" K $= \infty$ cannot be accepted literally, it is based on a real effect. As the firm adds capital and the probability of insolvency becomes negligible, it essentially acts as a financial intermediary. The cost of holding additional capital becomes correspondingly negligible, or at least no greater than for other non-risk-bearing intermediaries. The policyholders, of a firm that chooses $K = \infty$ receive complete protection against insolvency.

In considering joint equilibria with respect to K, P, and Q, we let P^* and Q^* represent the solutions corresponding to unlimited liability. Since $\rho = 0$ by assumption, P^* and Q^* are the solutions to

$$\mathrm{MR} = E_C/(1 + R_f) \tag{70}$$

and we let ξ^* denote the corresponding value for ξ:

$$P^*(1 + R_f) = E_C + \xi^* \bar{S}_C. \tag{71}$$

There are three cases to consider:

Case 1 $\psi(\xi^*) < R_f$.

Case 2 $\psi(\xi^*) > R_f$, and the unregulated firm will choose to set $K = \infty$.

Case 3 $\psi(\xi^*) > R_f$, and the unregulated firm will choose to set $K = 0$.

In case 1 the unregulated firm will choose to set $K = 0$ and will choose a corresponding P^{**} and Q^{**}. With the requirement $K \geq \bar{K}$ imposed on it, the firm will be led to set $K = \bar{K}$, raise its premium, and lower the number of policies written. As $\bar{K}$ is increased the probability of insolvency continues to fall, but it remains positive.

In case 2 regulation is obviously unnecessary. The unregulated firm chooses to be perfectly safe. Although monitoring costs exist, imposing a minimum capital requirement does not create costs of adjustment. In particular, imposing the requirement has no effect on P and Q.

In case 3 the firm behaves as in case 1 for low $\bar{K}$, but at a certain level for $\bar{K}$ the firm responds discountinuously and switches over to $K = \infty$. Sufficiently high minimum capital requirements will induce the firm to forsake the advantage of limited liability and add capital to the point where it is essentially perfectly safe.

These results suggest a schizophrenic regulatory environment. At least within the context of an insensitive demand curve, some firms are led to pursue an ultrasafe strategy. In that case solvency regulation, in the form of minimum levels for K, is unnecessary. But the results also imply that such regulation is costless, at least in the sense that it would not alter firm behavior, and thus inefficiencies will not arise as the firm adapts. For firms that would choose to set $K = 0$, regulatory constraints on minimum K will obviously have an effect. In some cases a firm that would choose $K = 0$ in the absence of regulation will choose to set $K = \infty$ when required to set $K \geq \bar{K}$!

What do the results imply for regulatory policy? Just as the results are somewhat paradoxical, so the implications may depend on how one views the different conclusions we derive.

"Completely safe" firms may arise naturally in a context in which applicants have only primitive notions of how to discriminate among firms, and use simple rules to try to attain safety. Not only are astute applicants not needed for safe firms to arise; indeed, it is the existence of applicants who rely on simple rules that creates the intangible capital that encourages the firm to protect its existence.

Thus, one can argue that regulation may not be needed in exactly the situation where it appears most clearly justified: where applicants are largely oblivious of the details of firms' financial structures and rely on simple proxies to provide them with safety. Suppose applicants generally divide firms into old or established firms and new or unknown firms. If applicants treat "old" firms as "safe" and are willing to pay a premium for safety, this creates the intangible capital that leads the firm to want to protect its position and to set $K = \infty$. The belief that old firms are safe becomes self-fulfilling.

If there a sufficient supply of firms viewed as old, the premiums of old firms will still be driven down to a competitive level but the competitive level will reflect the full costs of effectively unlimited liability. Those applicants who want safety can buy safety, and the market will provide it. Those applicants who are willing to rely on their own judgment and who may be willing to accept some nontrivial risk of insolvency are free to go to new or unknown firms. As long as all applicants understand that they have a choice of buying from an old firm or of relying on their own evaluations of the safety of a new firm, the case for regulation seems to degenerate to ensuring the protection of applicants who put more faith in their ability to judge firms than is warranted. Solvency regulation, in this view, protects not the naive but the arrogant.

But this conclusion assumes that the market position of each of the old firms is sufficiently profitable to protect by setting $K = \infty$. If the profitability of an old firm is too low, to the point where the incentives of limited liability come to dominate the value of protecting its position, then those applicants who are willing to pay for safety will be unwittingly exploited.

What conditions might lead to the profitability of old firms dropping to the point where it is not worthwhile to protect the intangible capital inherent in the perception that the firm is old and reliable? It is important to note that the threat of excess capacity is essentially absent; the amount of specific physical capital in the insurance industry is negligible. Marginal and average costs vary little, and firms can contract from an level of extended "production" with few problems. Indeed, the ease with which insurance firms can contract has produced some of the dissatisfaction with the insurance market's performance in recent years.

Clearly, rate regulation is one possibility. Forcing rates below long-run full cost levels could eliminate the desire to protect one's market access. Indeed, firms have been required, against their desires, to stay in some insurance markets in some states. The leverage—for better or ill—exists because these have been multiline firms for which the insurance line at issue represents only a proportion of their business, and because only one or a few states were exploiting this leverage.

Similarly, while technical change and structural innovation can disrupt the position of "old" firms, the insurance product is basically fixed, although methods of packaging and selling the product undergo significant changes. But methods differ considerably across different lines—particularly between personal and commercial lines—and structural changes have tended to affect old, established firms in only one or a few lines at any time.

Thus, to the extent that "old" firms, which sell (at least in part) to those who seek essentially complete safety, are highly diversified across different lines, the likelihood that conditions in different lines would simultaneously deteriorate to the point where an insolvency would dominate the selective exit from some markets seems very low.[45] Obviously, if several state regulators were to keep rates below full-cost levels in several lines this conclusion would be in jeopardy.

Another factor that destroys this line of argument is the existence of guaranty funds, which reimburse those who have claims against insolvent firms. If applicants are aware of guaranty funds and believe they provide perfect protection against a partial claim recovery, all reason for concern with their insurer's financial condition disappears. Indeed, applicants will not simply be indifferent between firms with different financial policies; those firms that set $K = 0$ will be able to sell at a lower premium than those setting $K = \infty$. In both cases policyholders will have complete protection. The existence of a perfect guaranty fund allows policyholders to share in the benefits of limited liability without incurring any of the costs.

It would seem that the existence of a perfect[46] and understood guaranty fund should destroy an insurance market. Each firm would choose $K = 0$. Applicants would choose the lowest premium available. Thus, as applicants were added, their full expected costs would be added to the industry's expected costs but not to the firms underwriting them. As firms became insolvent the liabilities of surviving firms would grow, and presumably the expected costs of new entrants would increase as the unpaid liabilities grew. A type of externality would exist, in that a firm would impose expected costs on its competitors as it wrote more applicants.

The inevitable result seems clearly pathological. As firms become insolvent, their liabilities are spread over existing firms in proportion to market share. It seems possible that a point would be quickly reached at which new entrants would wait until all surviving firms had gone under.

Thus, the existence of a guaranty fund creates the need for financial regulation. However, it should be noted that financial regulation only modifies the problem created when applicants have no incentive to avoid financially risky firms. According to the model we have analyzed, firms will choose to set $K = \bar{K}$, the probability of insolvency will be positive, and the "externalities" problem will still exist.

Although the existence of a guaranty fund destroys the rationale for protecting a firm's existence which we have modeled here, other factors also generate intangible capital. All firms incur startup costs and fixed costs in becoming established. Furthermore, writing an individual policy has fixed costs which are recouped only as the policyholder continues to renew. Thus, being an established firm has value even if applicants are indifferent to all firm attributes except premium rate. A sufficiently large capital requirement will produce the case where the value of survival dominates the advantages of limited liability, and firms will choose to become perfectly safe in the face of financial regulations.

Thus, the combination of a guaranty fund and financial regulations may produce a viable insurance market in which policyholders have perfect protection against insolvency. However, this result may have substantial costs, particularly as it affects the vigor of competition in the market. Firms which service those willing to trade off premium costs against the risk of partial claims recovery are eliminated. If all firms were equally efficient and if competition among "old" firms were vigorous, this cost would seem likely to be slight. The lower premium offered by some firms, such as in cases 2 and 3 above, would be lower only because limited liability lowers expected costs. For a risk-averse individual who buys insurance to provide compensation in the event of some catastrophe, it seems evident that a lower premium would not be attractive if the cost saving related to some probability of a partial or zero claim payment.

Suppose, however, that competition among "old" firms is not as vigorous as it could be, and that in a world of perfect ability to judge firms' financial conditions premiums would be lower. Suppose further that old firms do not always innovate quickly as new insurance needs appear. Then some individuals, while risk-averse, will want to buy insurance from new firms—either because they have confidence in their ability to judge financial condition and have faith in certain new and unestablished firms, or because they have unusual insurance needs to which the established firms are not responding and are willing to accept some probability of insolvency in order to obtain a product more suited to their needs. Both situations appear to have existed in recent years.

In this case, the effect of financial regulations on entry and on the vigor of competition in the industry must be considered. As we show elsewhere (Munch and Smallwood 1980), the effects of solvency regulation seem to be clearest with respect to the total supply of firms in the industry rather than with respect to insolvency rates. As discussed in that paper, financial regulations offer opportunities for abuse by a regulator who accepts the position that competition in the industry should be stifled.

The danger that financial regulations may protect policyholders from competition rather than from unrecognized risks is particularly clear in commercial lines. When firms find it difficult to join and form their own insurance pools to cover product liability, or when groups of professionals cannot establish new facilities for providing malpractice coverage, there is obviously a question as to who benefits from the regulations. The necessity for financial regulations across all lines to ensure solvency, and the existence of guaranty funds, should be evaluated critically.

Appendix A

Let M be total market flows at the end of the period in the absence of the insurance firm in question. The value V_M of ownership of rights to M is

$$V_M = \frac{E(M) - \theta \operatorname{Cov}(M,M)}{1 + R_f} = \frac{E(M) - \theta \operatorname{Var}(M)}{1 + R_f}$$

according to the capital asset pricing model. We can treat the introduction of the firm in two steps. First, the additional cash flow represented by the additional firm:

$$\mu^* = (1 + R_f)(K + PQ) - \sum_{i=1}^{Q} C_i$$

is added to total end-of-period market flows:

$$U = M + \mu^*.$$

The value of θ and all asset values must adjust, and we have

$$V_U = \frac{E(U) - \theta^* \operatorname{Cov}(U,U)}{1 + R_f}$$

since an infinite supply of a risk-free asset with rate of return R_f is assumed. Thus, for a specific cash flow D_j, we have

$$V_{D_j} = \frac{E(D_j) - \theta^* \operatorname{Cov}(D_j, U)}{1 + R_f}.$$

The insurance firm now invests a proportion α of its funds in a balanced portfolio which represents the total market (including the firm in question). The net end-of-period cash flow arising from the investment is

$$\mu^{**} = \alpha(R_U - R_f)(K + PQ),$$

where $V_U(1 + R_U) = U$ by definition. Let d denote the fraction of the total market which the firm has purchased. Thus, we have

$$\alpha(K + PQ) = dV_U$$

at the beginning of the period, and

$$\alpha(1 + R_U)(K + PQ) = dU$$

at the end. Therefore,

$$\begin{aligned}\mu^{**} &= \alpha(1 + R_U)(K + PQ) - \alpha(1 + R_f)(K + PQ)\\ &= dU - d(1 + R_f)V_U,\end{aligned}$$

so that

$$E(\mu^{**}) = dE(U) - d(1 + R_f)V_U,$$

$$\operatorname{Cov}(\mu^{**}, U) = \operatorname{Cov}(dU, U) = d\operatorname{Cov}(U, U).$$

Hence,

$$\begin{aligned}V^{**} &= \frac{dE(U) - d(1 + R_f)V_U - \theta^* d\operatorname{Cov}(U,U)}{1 + R_f}\\ &= d\left[\left(\frac{E(U) - \theta^* \operatorname{Cov}(U,U)}{1 + R_f}\right) - V_U\right]\\ &= 0.\end{aligned}$$

To compute V^*, the market value of the rights to μ^*, we use

$$E(\mu^*) = (1 + R_f)(K + PQ) - QE_C.$$

Since $(1 + R_f)(K + PQ)$ is constant, it follows that

$$\mathrm{Cov}(\mu^*, M+\mu^*) = \mathrm{Cov}\left(-\sum_{i=1}^{Q} C_i, M - \sum_{i=1}^{Q} C_i\right)$$

$$= -\sum_{i=1}^{q} \mathrm{Cov}(C_i, M) + \mathrm{Var}\left(\sum_{i=1}^{Q} C_i\right)$$

$$= -QV_M\sigma_{\mathrm{m},C} + Q\sigma_C + Q(Q-1)\gamma\sigma_C$$

because

$$\mathrm{Cov}(C_i, M) = \mathrm{Cov}(C_i, (1 + R_{\mathrm{m}})V_M)$$

$$= V_M \mathrm{Cov}(C_i, R_{\mathrm{m}}).$$

Appendix B

In the third section we derived

$$W(K) = \frac{(1 + R_{\mathrm{f}})[PQ(1 - F_\delta) - KF_\delta] - QE_\delta}{R_{\mathrm{f}} + F_\delta}.$$

From the definitions of $\delta(K)$, F_δ, and E_δ, it follows that

$$\frac{d}{dK}F_\delta = -\frac{f(\delta)}{Q},$$

$$\frac{d}{dK}E_\delta = \frac{\delta f(\delta)}{Q},$$

since only K is treated as variable in that section. If we multiply both sides above by $R_{\mathrm{f}} + F_\delta$ and differentiating with respect to K, it follows that

$$\frac{dW(K)}{dK} = \left(\frac{1}{R_{\mathrm{f}} + F_\delta}\right)\left[\left(\frac{W(K)}{Q}\right)f(\delta) - (1 + R_{\mathrm{f}})F_\delta\right]$$

$$= A(K)\left[\left(\frac{1}{1 + R_{\mathrm{f}}}\right)\left(\frac{W(K)}{Q}\right) - \frac{F_\delta}{f(\delta)}\right],$$

where

$$A(K) = \left(\frac{1 + R_{\mathrm{f}}}{R_{\mathrm{f}} + F_\delta}\right)f(\delta)$$

is a function of K that is always positive, if $f(\delta) > 0$ for all K.

For an average claim distribution that is a normal distribution, or approximately normal, the ratio $F_\delta/f(\delta)$ is monotonically decreasing. The class of distributions for which this ratio is monotonically decreasing are known as the increasing-failure-rate distributions in the literature of statistical reliability theory. These distributions, which include the gamma $G_{\lambda,\alpha}$ distributions with $\alpha > 1$, the truncated normal distribution, and the χ^2 distributions, provide approximately normal distributions which are non-negative and have thick upper tails, and are therefore appealing as a claims distribution for a firm with a finite number of policyholders. Thus, our results certainly do not depend on less appealing aspects of assuming a normal distribution.

Suppose that, at $K = K_1$, $W(K)$ has a positive slope. Then, since $W(K)/Q$ is growing while $F_\delta/f(\delta)$ is falling, the slope will continue to be positive for $K_2 > K_1$. Thus, it follows that

$$\left.\frac{dW}{dK}\right|_{K=K_1} > 0 \Rightarrow \left.\frac{dW}{dK}\right|_{K=K_2} > 0 \qquad \text{for } K_2 > K_1.$$

We now indicate how $W(K)/Q$ can be written in terms of standardized parameters, assuming that average claims $\bar{C}$ have a normal distribution

$$C \sim N(E_C,(\bar{S}_C)^2),$$

so

$$t = (\bar{C} - E_C)/\bar{S}_C \sim N(0,1).$$

For the normal distribution, we have

$$E_\delta = E_C(1 - \Gamma(Z)) - \bar{S}_C(\phi(Z)).$$

With the definition of $\delta(K)$, we have

$$W(K) = \frac{(1 + R_f)[PQ(1 - F_\delta) - KF_\delta] - QE_\delta}{R_f + F_\delta}$$

$$= \frac{(1 + R_f)PQ - \delta \cdot F_\delta \cdot Q - QE_\delta}{R_f + F_\delta},$$

so

$$\frac{W(K)}{Q} = \frac{(1 + R_f)P - \delta \cdot F_\delta - E_\delta}{R_f + F_\delta}.$$

Since

$$P(1 + R_f) = E_C + \xi \cdot \bar{S}_C,$$

we can use the expressions for E_C, $\bar{S}_C$, and $\delta(K)$ to obtain

$$\frac{W(K)}{Q} = \frac{[\xi - Z\Gamma(Z) + \phi(Z)]\bar{S}_C}{R_f + \Gamma(Z)}.$$

Since $Z(0) = \xi$, we have

$$\frac{W(0)}{Q} = \frac{[\xi - \xi\Gamma(\xi) + \phi(\xi)]\bar{S}_C}{R_f + \Gamma(\xi)}.$$

As K approaches infinity, $Z(K)$ also approaches infinity while $Z\Gamma(Z)$ and $\phi(Z)$ approach zero, so

$$\frac{W(\infty)}{Q} = \frac{\xi\bar{S}_C}{R_f},$$

from which it follows that

$$W(\infty) > W(0) \qquad \text{if } R_f < \xi\Gamma(\xi)/[\phi(\xi) - \xi\Gamma(\xi)].$$

Let Z correspond to $\delta(K)$, the value of average claims below which the firm is insolvent:

$$Z(K) = [\delta(K) - E_C]/\bar{S}_C.$$

If we let $\phi(\cdot)$ represent the density function for the standard normal, it follows that

$$f(\bar{C}) = \phi(t)/\bar{S}_C$$

and

$$dC = (\bar{S}_C)dt,$$

so

$$\begin{aligned} F_\delta &= \int_{\delta(K)}^{+\infty} f(\bar{C})d\bar{C} \\ &= \int_{[\delta(K)-E_C]/\bar{S}_C}^{+\infty} \phi(t)dt \\ &= \int_{Z(K)}^{+\infty} \phi(t)dt \\ &= \Gamma(Z), \end{aligned}$$

where $1 - \Gamma(Z)$ is the cumulative distribution function of a standard normal random variable. Similarly, integrating E_δ by parts leads to

$$E_\delta = \int_{-\infty}^{\delta(K)} \bar{C} f(\bar{C}) d\bar{C}$$

$$= E_C(1 - F_\delta) + (\bar{S}_C) \int_{-\infty}^{Z} t\,\phi(t)dt.$$

Because[47]

$$\int_{-\infty}^{Z} t\,\phi(t)dt = -\phi(Z),$$

we have

$$E_\delta = E_C(1 - F_\delta) - \bar{S}_C \phi(Z).$$

Last, we show that $P^{**} > P_1^{**}$, where P^{**} and P_1^{**} are the multiperiod and one-period solutions, respectively, for the case in which P and Q are allowed to vary, with K fixed at zero. Differentiating W with respect to Q produces

$$(R_f + F_\delta)\left(\frac{\partial W}{\partial Q}\right) + \left(\frac{\partial F_\delta}{\partial Q}\right) W = \left(\frac{\partial (PQ)}{\partial Q}\right)(1 + R_f)(1 - F_\delta) - E_\delta,$$

because

$$\left(\frac{\partial E_\delta}{\partial Q}\right) = -P(1 + R_f)\left(\frac{\partial F_\delta}{\partial Q}\right).$$

Since P^{**} occurs at a point at which

$$\frac{\partial W}{\partial Q} = 0$$

and since $W > 0$ and

$$\frac{\partial F_\delta}{\partial Q} > 0,$$

it follows that P^{**} occurs at a point where

$$(1 + R_f)\left(\frac{\partial (PQ)}{\partial Q}\right) > \frac{E_\delta}{1 - F_\delta} > E_\delta$$

and thus that $P^{**} > P_1^{**}$, if marginal revenue is monotonically decreasing.

Appendix C

Let $C_1,\ldots,C_Q$ be identically distributed with a common mean μ_C. We conceptualize a model in which μ_C is first determined randomly, which then defines the joint distribution of the C_j. Let E_μ refer to expectations taken with respect to the distribution that produces μ_C, let $E_{C/\mu}$ refer to the conditional expectations of the C_j given μ_C, and let E refer to the unconditional expectations which reflect the combined stochastic processes. Then μ_C has first moment

$$E_\mu(\mu_C) = \bar{\mu}_C$$

and second moment

$$\mathrm{Var}_\mu(\mu_C) = v^2,$$

while the conditional moments of the C_j are

$$E_{C/\mu}(C_j/\mu_C) = \mu_C,$$

$$\mathrm{Var}_{C/\mu}(C_j/\mu_C) = \sigma_C,$$

$$\mathrm{Cov}_{C/\mu}(C_i,C_j/\mu_C) = \gamma\sigma_C,$$

$$\mathrm{Cov}_{C/\mu}(C_i,R_m/\mu_C) = \rho S_m S_C.$$

Our purpose is to compute the unconditional moments, which show that we can make the model implicitly reflect "uncertainly about μ_C" by appropriately specifying the assumed parameters. The mean μ_C is assumed to be distributed independently of both the C_j and R_M. Letting * denote unconditional moments, we have

$$\begin{aligned}
\mu_C^* &= E(C_j) \\
&= E_{C/\mu}\{E_\mu(C_i)\} \\
&= E_\mu\{E_{C/\mu}(C_i)\} \\
&= E_\mu(\mu_C) \\
&= \bar{\mu}_C,
\end{aligned}$$

$$\begin{aligned}
\sigma_C^* &= E[(C_i - \bar{\mu}_C)^2] \\
&= E[(C_i - \mu_C + \mu_C - \bar{\mu}_C)^2] \\
&= E_\mu\{E_{C/\mu}[(C_i - \mu_C)^2 + 2(C_i - \mu_C)(\mu_C - \bar{\mu}_C) + (\mu_C - \bar{\mu}_C)^2]^2\}
\end{aligned}$$

$$= E_\mu[\sigma_C + (\mu_C - \bar{\mu}_C)^2]$$
$$= \sigma_C + E_\mu[(\mu_C - \bar{\mu}_C)^2]$$
$$= \sigma_C + \upsilon^2.$$

It similarly follows that

$$\mathrm{Cov}^*(C_i, C_j) = \mathrm{Cov}(C_i, C_j) + \upsilon^2,$$

$$\mathrm{Cov}^*(C_i, R_\mathrm{m}) = \mathrm{Cov}(C_i, R_\mathrm{m}).$$

Thus, we can write

$$\mu_C^* = \bar{\mu}_C,$$

$$\sigma_C^* = \sigma_C + \upsilon^2,$$

$$\mathrm{Cov}^*(C_i, C_j) = \delta^* \sigma_C^* = \mathrm{Cov}(C_i, C_j) + \upsilon^2 = \delta\sigma_C + \upsilon^2,$$

$$\mathrm{Cov}^*(C_i, R_\mathrm{m}) = S_C^* S_\mathrm{m} \rho^* = \mathrm{Cov}(C_i, R_\mathrm{m}) = S_C S_\mathrm{m} \rho.$$

Therefore,

$$\delta^* = \frac{\delta\sigma_C + \upsilon^2}{\sigma_C^*}$$
$$= \frac{\delta\sigma_C + \upsilon^2}{\sigma_C + \upsilon^2}$$
$$= \delta\left(\frac{\sigma_C}{\sigma_C^*}\right) + \left(\frac{\upsilon^2}{\sigma_C^*}\right)$$
$$= \delta\left(\frac{\sigma_C}{\sigma_C + \upsilon^2}\right) + 1\left(\frac{\upsilon^2}{\sigma_C + \upsilon^2}\right),$$

while

$$S_C^* S_\mathrm{m} \rho^* = S_C S_\mathrm{m} \rho,$$

$$\rho^* = \left(\frac{S_C}{S_C^*}\right)\rho.$$

Thus, if there is uncertainty about the true mean, the effect on the relevant moments is that the claims variance σ_C is increased ($\sigma_C^* > \sigma_C$) by an amount equal to the variance of the uncertain mean, and the pairwise correlation of claims is increased ($\delta^* > \delta$; note that δ^* is a weighted average of δ and 1), while the correlation between each claim and the market rate of return is decreased in absolute value ($|\rho^*| < |\rho|$, since $S_C^* > S_C$).

In particular, note that claims may be uncorrelated, given the mean ($\delta = 0$), but that uncertainty about the mean of the claims distribution ($v^2 > 0$) implies that the overall correlation among claims is positive ($\delta^* > 0$ if $\delta = 0$) and is then equal to the ratio of the variance of the mean to the total unconditional variance.

Notes

1. The McKinsey study of 101 insolvencies of life companies and 129 insolvencies of property-liability companies found dishonesty to be the primary cause of insolvency much more frequently for life companies (77 percent) than for property-liability companies (34 percent). The principal cause (59 percent) of property-liability insolvencies has been underwriting losses.

2. The firm is not able to distinguish between better and poorer risks in the relevant population, and thus there is no active policyholder selection.

3. The "riskless asset" may represent bonds with essentially zero default risk, such as government bonds. The impossibility of hedging perfectly against inflation is ignored.

4. This conclusion depends on the assumption of zero transactions costs in the securities market. We comment on the relevance of the CAPM framework below.

5. Throughout the analysis we ignore the effect of the firm in question on the distribution of R_m. However, we do not ignore its impact on total flows of funds in the market at the end of the period, as will be clear below.

6. The correlation between the market asset and claims costs can reflect various phenomena, such as the effect of unanticipated inflation on both claims and the market asset. Similarly, it is widely believed that OPEC price increases affected both the stock market and gasoline consumption and thus automobile insurance claims.

7. That is, to the extent that variations in μ^+ are correlated with U, they cannot be effectively "diversified away" as all investors devote an insignificant proportion of their portfolios to the ownership of this firm.

8. Note that assumed knowledge must include parameters which reflect the firm's uncertainty about the location of the claims distribution.

9. Furthermore, the results depend critically on assumptions made about the nature of the partial ignorance, about the awareness of buyers of their limitations and their response to it, and about the behavior of firms in such a context and their ability or willingness to exploit buyer behavior.

10. If the returns from advertising are not entirely exhausted in the current period, a firm that advertises extensively can be assumed not to expect immediate dissolution, *ceteris paribus*.

11. Indeed, we show below that such behavior tends to be self-fulfulling: To the extent that demand is insensitive to current choices, firms will have the incentive to become very safe.

12. The capital asset pricing model implicitly presumes an infinitely elastic supply of the risk-free security; our inclusion of K as a cash flow thus has no effect on the conclusions.

Treating μ^* as a new flow into the market assumes that the writings of other insurance firms remain unchanged as the firm in question changes its premium rate.

13. The parameter λ is often defined alternatively as $[E(R_U) - R_f]/S_U$, with a corresponding adjustment in the equilibrium equation.

14. "Rate of Return & Profit Provision in Automobile Insurance," State Rating Bureau, Division of Insurance.

15. Assuming that the firm chooses to stay in business; we discuss this possibility briefly below.

16. Also involved is the implicit assumption that after insolvency the firm cannot be resurrected and write policies according to the given demand curve.

17. Of course, the firm's past experience will affect the investment environment of its owners by changing their wealth; we ignore this effect.

18. It is clear that the overall optimal policy for the firm is a stationary policy.

19. The CAPM is essentially a one-period model. However, it can be applied by noting that Y represents the net cash flow of an owner who sells his interest in the firm at the end of the period, after claims have occurred but before paid-in capital has been restored.

20. Equation (41) could alternatively be obtained by assuming that all investors are risk-neutral. However, the CAPM conclusions are obtained on the assumption that investors are risk-averse, to an unspecified degree. Equation (41) follows not from risk-neutrality, but rather from the assumption that investors can costlessly diversify their own portfolios.

21. In analyzing the behavior of $W(K)$, we treat Q as fixed, ignoring the dependence of $Q(K)$. This assumption is relaxed in the fourth section.

22. The product $K \cdot F_\delta$ approaches zero for any distribution.

23. This conclusion can be derived for the increasing failure rate distributions, which include both the truncated normal and other approximately normal distributions that are non-negative and have thick upper tails, and are thus appealing as assumed claims distributions for a firm with a finite number of policyholders.

24. For simplicity of exposition, we ignore the case $W(0) = W(\infty)$, in which the firm is perfectly indifferent between $K = 0$ and $K = \infty$.

25. Of course, the change in $V(K) - K$ becomes trivially small as K becomes large.

26. The distribution of $\tilde{C}$ must approach a normal distribution as Q increases, by the central limit theorem.

27. In terms of the previously defined parameters, $(\bar{S}_C)^2 = [1 + (Q - 1)\gamma](S_C)^2/Q$.

28. $\psi(\xi) = \xi Z(\xi)/[\phi(\xi) - \xi Z(\xi)]$, where $\phi(\xi)$ is the density function of the standard normal distribution and $Z(\xi)$ is the probability that a standard normal random variable is greater than ξ.

29. R_f represents the real risk-free rate of return.

30. The case $\gamma = 0.10$ can be interpreted as a claims environment in which 10 percent of the total variance in the distribution of individual policyholder claims is attributable to uncertainty about the location of the distribution (that is, uncertainty about the location of the mean) and 90 percent represents the variation of claims about the true mean.

31. The range includes $E_C = 100.0$; $\bar{S}_C = 20.0, 30.0$; $R_f = 0.03, 0.06, 0.10, 0.15, 0.20$; $\xi = 0, 0.025, 0.050, 0.075, 0.10, 0.15, 0.20, 0.50, 1.0, 2.0$; $b = 0, 0.05, 0.10, 0.15, \ldots, 0.95, 1.0$.

32. However, the specific values of ξ for which the different cases appear vary.

33. Although Q (the number of policyholders), does not appear explicitly, it affects $\bar{S}_C$.

34. That is, it was assumed both that $\alpha = 0$ and that $\rho = 0$; these assumptions continue to be maintained in this section.

35. As in the third section, the firm's owners clearly will not wish to maximize total market value $V(P,Q,K)$ with K variable, but net market value $W(P,Q,K) = V(P,Q,K) - K$.

36. It may be that if P, Q, and K are allowed to vary simultaneously, a still better position could be found. The combination $[P,Q,K]$ that represents the best possible simultaneous choice of all three variables is the global optimum.

37. The combination $[P^{**},Q^{**},K{=}0]$ dominates $[P^*,Q^*,K{=}0]$ (because P^{**} and Q^{**} are optimal with $K = 0$), which dominates $[P^*,C^*,K{=}\infty]$ (because $\psi(\xi^*) < R_f$), which dominates $[P^{**},Q^{**},K{=}\infty]$ (because P^* and Q^* are optimal with $K = \infty$). Thus, if $[P^{**},Q^{**},K{=}\infty]$ dominated $[P^{**},Q^{**},K{=}0]$, one would have a mutually inconsistent chain, as in an Escher drawing.

38. Marginal expected policyholder cost is the mean of the claims cost distribution that has been truncated at the point at which insolvency occurs. As Q is increased, the truncation point falls.

39. Thus, E_C^∞ corresponds to E_C.

40. One can prove that ξ falls monotonically as Q increases over the relevant range. Multiplication of equation (54) by Q, differentiation with respect to Q, and manipulation leads to

$$\frac{\partial \xi}{\partial Q} = \frac{(1 + R_f)MR - E_C}{S_C\sqrt{Q}} - \frac{\xi}{2Q}$$

for the case $\gamma = 0$. It is simple to then show that monotonicity continues to hold for $\gamma > 0$.

41. Proof that $P^{**} > P_1^{**}$ is given in appendix B. However, it is not obvious that P^{**} will be lower than P^*. Intuitively, it seems that the influence of limited liability, which reduced the optimal premium rate, should dominate the influence of the survival motive; after all, we are considering the case where, at the premium rate P^*, the firm chooses to set $K = 0$. But we have no proof for this conjecture.

42. (Because premiums are received at the beginning of the period while claims are paid at the end.)

43. Given the assumption $\rho = 0$, the CAPM equation for market value collapses down to the sum of discounted expected net profits, adjusted for the probabilities of survival. As above, this result is not based on an assumption that investors are risk-neutral, but rather reflects their opportunities for portfolio diversification.

44. The conclusions that the optimal K is either $K = 0$ or $K = \infty$ is obtained by assuming that average claims costs follow an increasing risk probability distribution, which includes both the normal and related distributions that span the range of plausible distributions. The characterization (69) assumes that the distribution of average claims costs is normal.

45. Geico is highly specialized to automobile insurance.

46. As opposed to a fund that provides only partial compensation, or provides compensation with a significant lag, or with significant associated hassle.

47. We thank Gus Haggstrom for pointing this out.

References

Hammond, J. D., Shapiro, Arnold F., and Shilling, Ned. 1978. Regulation of Insurer Solidity through Capital and Surplus Requirements. Technical report, Department of Insurance, Pennsylvania State University.

Hofflander, Alfred E. 1969. "Minimum Capital and Surplus Requirements for Multiple Line Insurance Companies: A New Approach." In S. L. Kimball and H. S. Denenberg (eds.), *Insurance, Government, and Social Policy* (Homewood, Ill.: Irwin).

Jensen, Michael C. 1972. "Capital Markets; Theory and Evidence." *Bell Journal of Economics and Management Science* 3: 357–398.

Joskow, Paul. 1973. "Cartels, Competition, and Regulation in the Property/Liability Insurance Industry. *Bell Journal of Economics and Management Science* 4: 375–427.

Mayerson, Allen L. 1969. "Ensuring the Solvency of Property and Liability Insurance Companies." In S. L. Kimball and H. S. Denenberg (eds.), *Insurance, Government, and Social Policy* (Homewood, Ill.: Irwin).

McKinsey and Co., Inc. 1974. Strengthening the Surveillance System. Report to National Association of Insurance Commissioners.

Munch, P., and Smallwood, D. 1980. "Solvency Regulation in the Property/Casualty Insurance Industry: Empirical Evidence." *Bell Journal of Economics* 11: 261–279.

National Association of Insurance Commissioners. 1974. Monitoring Competition. Report.

Olson, Douglas, G. 1970. Insolvencies among Automobile Insurers. Report to U.S. Department of Transportation.

Scott, J. H., Jr. 1976. "A Theory of Optimal Capital Structure." *Bell Journal of Economics* 7: 33–54.

Comment

Howard Kunreuther

In this interesting and stimulating paper Munch and Smallwood raise the following question: Does solvency regulation of the casualty insurance industry produce sufficient benefits in the form of consumer and owner protection to justify the operating and monitoring costs of such a system?

The authors appropriately note that consumers may be imperfectly informed about the financial stability of different companies. Two empirical studies support their point by shedding additional light on the imperfect information consumers have been obtaining on insurance. J. D. Cummins et al. ("Consumer Attitudes Toward Auto and Homeowners Insurance," Wharton School Department of Insurance report) found in a field survey of 2,462 individuals who purchased automobile and/or homeowners coverage that few compared the policy terms from different companies before making their purchase decision. Similar behavior was found by Kunreuther et al. (*Disaster Insurance Protection: Public Policy Lessons* [New York: Wiley-Interscience, 1978]) in a field survey of 3,000 homeowners residing in flood- and earthquake-prone areas. This study revealed that most homeowners were either unaware of terms in their policies or had misinformation on such relevant data as premiums and deductibles. If consumers are reluctant to collect data on these characteristics, which directly affect them, it is unlikely that they have information on the financial quality of the company. The latter information would require detailed reading of balance sheets and perhaps inside information on recent company performance and strategy.

The above empirical findings raise a question of whether the capital assets pricing model (CAPM) is an appropriate model for describing an insurance firm's behavior. If companies know that consumers have limited interest in collecting detailed information on insurance and hence have imperfect data on policy terms, then their behavior and marketing strategies are likely to differ from actions they would take if they assumed that consumers collected detailed information and expected them to try to maximize utility. There is no way to answer specifically the question of what assumptions firms are making about consumer behavior unless we undertake more detailed empirical analyses. However, on a theoretical plane it should be possible to investigate a firm's profit-maximizing behavior by postulating that the consumer has limited and imperfect information on characteristics of the insurance market.

The interesting work Munch and Smallwood have undertaken on the CAPM can be used as a starting point for determining whether alternative models produce significantly different results in insurance companies' behavior. In particular, one should then be able to discover whether different assumptions about consumer behavior will affect the actions of firms in a way that influences bankruptcy possibilities.

Let me turn now to the model Munch and Smallwood investigate. To make their multiperiod analysis tractable from a mathematical point of view, the authors make the following simplified assumptions regarding consumer and firm behavior:

- that consumers are indifferent to the financial condition of the firm, either because they have imperfect information, because they are protected through a guaranty fund, or because they buy insurance solely to satisfy responsibility requirements; and
- that firms do not invest in risky market securities.

As a result of these assumptions, a firm should choose to set its financial capital (K) equal to either 0 or ∞. The choice of either extreme is determined by the relative advantages of minimizing the liability should the firm go bankrupt ($K = 0$) and of reducing the probability of bankruptcy to a negligible level ($K = \infty$). The reason for this dichotomous policy is that the firm, if it exists, faces the same demand curve and claims policy as in the previous period.

To make the model more realistic and interesting, Munch and Smallwood might want to view consumer demand as a function of both the premium and the perceived financial stability of the firm as represented by K. One way to do this in the spirit of their analysis is to divide consumers into old customers (who continue to purchase a policy from the firm with which they started, independent of its financial stability) and new customers (who may be sensitive to the financial condition of firms in making their choice, but can misperceive this information). In each period, a fraction of old customers leave the market and another fraction of new customers enter. In this case the firm will have to concern itself with the impact that K has on demand so that the optimal value may be somewhere between 0 and ∞. This paper discusses a special case of this more general model: New customers do not exist, so demand is solely a function of the premium (P).

The performance of several other regulatory systems, aside from minimum capital requirements, could be investigated by the authors in the context of the broader model proposed above. They might want to investigate the performance of a policy where K is a function of the number of policyholders (Q). Such a policy may be more desirable than specification of a fixed K if the variance of the claims distribution, and hence the probability of bankruptcy, were assumed to change with Q.

Another alternative the authors might want to investigate would be to have the regulatory commissioner specify a minimum level of capital, K, below which a firm would have to undertake special steps to avoid bankruptcy. These could include raising additional capital from outside sources or reducing the number of policyholders in future periods (for example, by not renewing all policies). Under such a system, K serves the same function as safety stock in inventory systems where the firm faces uncertain future demand. The minimum capital requirement would then directly affect the future probability of bankruptcy by forcing the firm to pay special attention to factors that affect it directly.

As Munch and Smallwood indicated in an earlier version of this paper, a variant of this system is in operation in some states which require a minimum capital requirement (K) and an additional surplus requirement (S). Once a firm's assets fall below $K + S$, it is forced to take special steps such as adding to reserves. By developing a theoretical analysis of this problem, Munch and Smallwood could address the implications of such a system for firm behavior and insolvency probabilities. Admittedly there are monitoring problems associated with measuring surplus, as the authors pointed out, and these costs would have to be included in an evaluation of such a regulatory policy. A more detailed discussion can be found in A. L. Mayerson's "Ensuring the Solvency of Property and Liability Insurance Companies," in *Insurance, Government and Social Policy*, ed. S. L. Kimball and H. S. Denenberg (Homewood, Ill.: Irwin, 1969).

In an earlier version of this paper, the authors provided empirical evidence on insolvencies which indicates that failures are more common among firms writing automobile insurance than engaging in commercial lines. On the basis of this evidence the authors conclude that these data are consistent with the hypothesis that insolvencies may be caused by poorly informed purchasers of auto coverage.

Empirical evidence from a study by D. Olson (*Insolvencies Among Automobile Insurers* [Washington, D.C.: U.S. Government Printing Office, 1970]) suggests a different interpretation of why insolvencies have occurred in the automobile insurance market. In recent years many states have instituted financial responsibility laws and/or have required automobile insurance as a condition for a license. (A financial responsibility law requires a driver who has caused an accident to show proof of ability to pay the loss, either through an insurance policy or by posting bond.) These structural changes naturally expanded the demand for insurance. During this same period many leading companies felt that automobile rates were inadequate, and hence were reluctant to renew existing policies let alone to extend market coverage to new customers. This gap between supply and demand led many new firms to enter the automobile insurance market. Olson cites detailed empirical evidence indicating that most auto-insurer insolvencies were due to "intentional management ineptness bordering on fraudulent behavior rather than on impersonal market forces" (p. 43). If newly formed companies were intentionally engaging in such behavior, they would be more likely to enter states where there was no minimum capital requirement.

Another factor that may have caused insolvencies in the automobile market is the difficulty firms may have had in determining risks on which their premiums were based. Olson indicates that expenses for automobile claims are not restricted to a single year, because individuals may be able to collect today on injuries from accidents that occurred in past years. Hence, premiums may not be easily determined by looking at the past claims distribution. The problem of setting economically variable rates is exacerbated by state regulatory agencies which restrict proposed rate increases by companies.

The difficulty firms have in estimating future losses provides an additional reason for instituting minimum capital requirements. If firms are unable to determine the mean and variance of their future claims from past data, then their decisions on optimal premiums and policyholders may be based on incorrect information. Minimum capital requirements may then serve the useful function of encouraging the entry of larger firms, which may be in a better position than smaller companies to absorb unexpected losses.

The authors may also want to investigate the impact of a requirement proposed by P. Joskow ("Cartels, Competition, and Regulation in the Property/Liability Insurance Industry," *Bell Journal of Economics* 4: 275–326) that all firms be required to carry complete insurance against bankruptcy, with rates varying directly with their premium/capital ratio. Such a requirement should reinforce minimum capital requirements in protecting firms and consumers against their own actions based on imperfect information.

The Munch-Smallwood paper is an interesting first step in addressing policy questions related to the impact of financial regulation on insolvency. Before one advocates specific recommendations, further work should be undertaken to relate the decision processes of consumers and firms to the institutional arrangements currently in force. Such research may then lead to more definitive answers to the general question about the desirability of regulation posed at the beginning of the paper.

Comment

Michael P. Lynch

Regulations that are designed to "insure" insurance-company solvency may have an important impact on the entire insurance industry, yet they have received very little attention from economists from either a theoretical or an empirical point of view. Munch and Smallwood provide a welcome first step toward a theory of the economics of solvency regulation by asking why an unregulated market will not automatically produce the efficient level of solvency and by proposing a model to answer this question. Their model is designed to apply to the casualty-and-property segment of the insurance industry, but it is worth noting that solvency regulation figures prominently in the $30-billion-a-year life insurance business. For example, a well-known textbook (D. McGill, *Life Insurance*, revised edition [Homewood, Ill.: Irwin, 1976]) states the following (p. 776):

> The primary purpose of state insurance regulation is to maintain the solvency and financial soundness of the companies providing insurance protection. In states having large domestic insurers the amount of effort expended in the supervision of the insurers' affairs exceeds that involved in all other kinds of supervisory work combined.

This emphasis on solvency regulation exists despite the rarity of insolvencies among life insurance companies and the even greater rarity of consumers being hurt by insolvency. (Equity Funding was a case of defrauding stockholders, not policyholders.)

Solvency regulations that include minimum capital or surplus requirements, restrictions on portfolio composition, reserve requirements, and asset-valuation requirements may have an important impact on the level and structure of life insurance prices. Indeed, these regulations may be an important cause of the poor rate of return on the savings element offered by life insurance relative to banks, savings and loan associations, and other financial intermediaries. Though it is difficult to find consumers who have been hurt by firms becoming insolvent, it is not at all difficult to find consumers who have paid higher than necessary amounts for their insurance coverage from highly solvent firms. These considerations suggest that there is a tradeoff between increased solvency and increased prices and that, at least in life insurance, the optimal insolvency rate may be higher than the one currently observed.

Another problem worth exploring is the reasons for the regulations in the first place. Who demanded them? Consumers? Reformers? The industry itself? The first two seem unlikely. I will suggest that the industry itself demanded these regulations, and that the benefits derived are subtle and cannot be measured in terms of reduced insolvencies.

Let me turn now to the theoretical analysis in the Munch-Smallwood article. As a first step in analyzing the costs and benefits of regulation, the authors quite appropriately ask how the probability of insolvency would be determined in an unregulated market. The owners of an insurance company, as well as its policyholders, have an interest in the company's solvency. Under some circumstances, at least, there may be no need for solvency regulation, since the owners will have the proper incentives to choose the "optimal" probability of solvency. The capital assets pricing model (CAPM), at first glance, appears to provide an attractive framework for analyzing this problem. It seems as though this model can be used to solve for the amount of paid-in capital; the number of policies to be sold; and the proportion of the firm's assets to be invested in risky assets rather than in the (elusive) riskless asset, which maximize the market value of the firm. Since solvency regulations commonly specify some minimum paid-in capital and impose restrictions on how much of a company's assets may be invested in certain "risky" assets, the CAPM seems to provide a convenient framework for the analysis. I believe that, in this case, appearences are deceptive.

The CAPM is a theory of how a given set of risk-averse investors value a given set of risky earning streams. It assumes that investors have full information on the probability distributions of these risky earnings streams and that they can and will bear unlimited liability for them. It is basically a theory of the demand for risky assets. Equilibrium prices are obtained by assuming that there is a fixed supply of risky assets, which must be held. The model is not useful for the problem at hand, for the following reasons:

- It focuses on the behavior of the investors in an insurance company, whereas the focus should be on the policyholders. It is the possibility that the insurance company may not pay valid claims, and the impact of this possibility on the demand for its product, that makes insurance-company insolvency special. As Munch and Smallwood point out, the unlimited-liability assumption rules out the problem of insolvency from the policyholders' point of view. They attempt to drop this assumption, but at the cost of assuming that investors are risk-neutral and that policyholders are indifferent to the level of paid-in capital. These assumptions appear to me to rule out any meaningful analysis of policyholder concern about insolvency.

• Contrary to appearances, the basic CAPM does not have any interesting implications for the optimal level of paid-in capital or the optimal mix of risky and riskless assets. The model determines a value for each asset. The current value of K dollars is simply K dollars. The current value of each insurance policy, the random variable $P - C_i$, is the discounted value (at the riskless rate of return) of the expected difference between the premium and the claim cost minus the market price of risk times the undiversifible risk between the policy and the total set of risky assets (including the policy itself). The market value of the "company" is simply the sum of the value of its assets. The value of the "firm" appears to be independent of the particular owners' choice of $\propto$, just as the value of the CAPM "firm" is independent of the owner's choice of a debt-equity ratio (the famous proposition 1 of F. Modigliani and M. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review* XLVIII [1958]: 261–297; for a proof that this proposition holds in the CAPM see J. Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics* XLVII [1965]: 13–37). The risk preferences of the initial owners influence the value of the firm only insofar as they influence the market price of risk. In equilibrium, all investors who hold any risky assets at all hold some risky insurance assets. That is, all risky investors are "owners" of the insurance firm.

• As far as determining the optimal number of policyholders, the Munch-Smallwood analysis is at least incomplete and perhaps mistaken. Every time a new policy is underwritten, a new random variable is added on the supply side of the market. This requires that a new equilibrium set of values be determined. This may result in a change in the equilibrium value of the "market price of risk" and a change in the total value of the universe of risky assets in the market. Munch and Smallwood implicitly assume that the market price of risk is independent of the number of policies written. Were this so, then the market value of each new policy would be the same as each old policy, at least in the special case where the insurance claim variables are assumed to be independent of all other risky assets in the market. If it was worthwhile to write one policy, then it would be worthwhile to write an indefinite number of policies; that is, the market value of the firm would increase without limit as the number of policyholders increased.

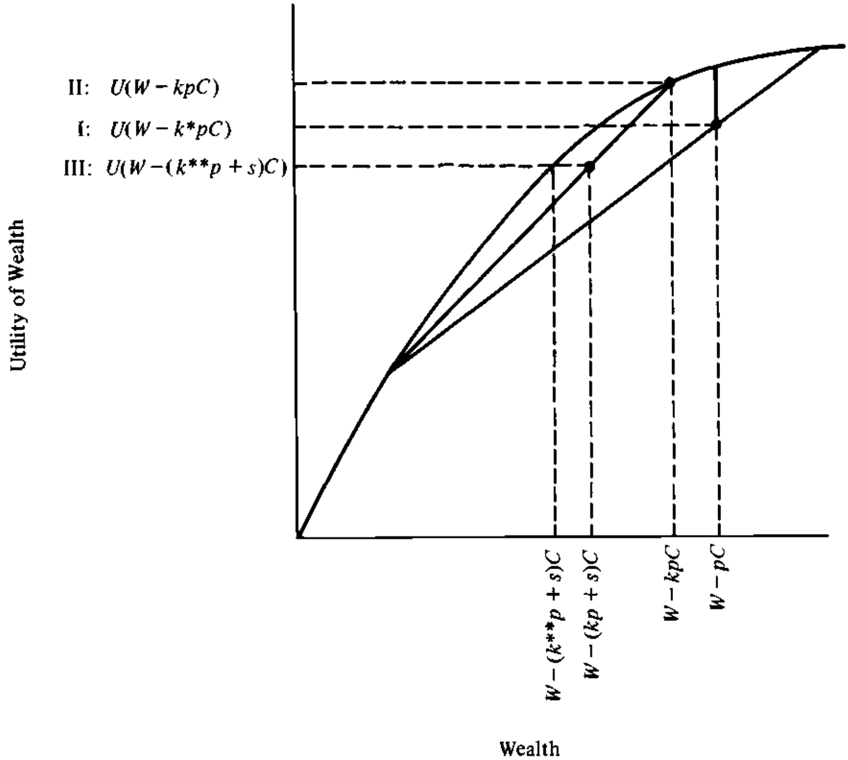
• As may be obvious from the above, I do not think the model has "interesting implications" concerning the proper treatment of investment income for insurance ratemaking purposes. The complexities of the investment income problem arise more from the difficulty of determining how much of the income derives from policyholders' capital rather than investors' capital than from determining the proper risk-adjusted rate of return for stockholders.

The CAPM does not seem to provide a useful framework for the analysis of insurance company solvency. I shall sketch an alternative model that may prove helpful. I start from the assumption that it is not the problem of the investors that makes an insurance company's possible insolvency special—indeed the problem doesn't arise for a large portion of the business which is sold through mutual companies. It is the effect of the possibility of insolvency on the policyholder that makes insurance company insolvency special and has led to the special regulations. The belief in the mind of a potential policyholder that the company may fail fundamentally alters the product that the company is trying to sell; that is, the fear of insolvency may greatly change the demand for the product.

Let us take the simplest case. A risk-averse individual is subject to a loss of C with probability p. He can purchase a full-coverage insurance policy for a premium kpC, where $k > 1$ and represents the "loading" in the policy. The figure portrays how this individual can be made better off so long as he can obtain an insurance policy with a loading between 1 and k^*, where the latter depends on his degree of risk aversion.

The situation changes radically if the consumer believes there is some probability s that the firm will be unable to pay off on a claim. The consumer can no longer pay a small amount to rid himself of large uncertain loss. He is still subject to the uncertain large loss, and, in addition, to the certain small premium payment. Graphically (see figure), this means that he trades a point on one line segment (such as I) for another point on another segment (for example, III) instead of trading a risky situation (I) for a riskless situation (II). This means that, for any given loading, the insurance is worth less to the individual than in the no-insolvency case, and for any given k makes it more likely that he will actually be made worse off if he buys the policy.

Suppose, as seems likely, that the consumers cannot judge the level of s very well, nor can they compare one company's chance of insolvency to that of another. The thing that is easy to compare is the premium that each company will charge to assume the risk. With other things equal, the lower the "loading" the greater the probability of insolvency. If consumers cannot distinguish one company from another in terms of solvency, and so choose their policies on the basis of premium alone, then the insolvency-prone companies will drive out the more solid



Impact of insolvency on demand for insurance coverage.

companies. Of course, as consumers learn that there is a significant chance of insolvency they will simply stop buying insurance policies. Thus, the end result is the destruction of the entire market. This story is similar, if not identical, to George Akerlof's "market for lemons" ("The Market for Lemons: Quality, Uncertainty, and the Market Mechanism," *Quarterly Journal of Economics* LXXIV [1970]: 488–500).

From this point of view, the benefits of solvency regulation cannot be measured merely in terms of reduced insolvency rates. The main benefits are in the increased size of the market which is made possible by the policyholders' belief that insolvencies either won't occur or that, if they do, the policyholders will not be hurt.

I have not made a detailed study of the history of life-insurance solvency regulations, but what little I know of it is consistent with the "lemon" theory of insolvency regulation. In the 1840s the life insurance industry began to grow very rapidly with the successful introduction of the "mutual" policy and the beginnings of the agency system (see J. O.

Stalson, *Marketing Life Insurance: Its History in America*, revised edition [Homewood, Ill.: Irwin, 1969], pp. 217–236 and 292–326). Success attracted new entrants, some of whom began to offer "dividends" (paid in scrip, not cash) amounting to 70–80 percent of the annual premium. Agents for company A would suggest that company B was offering dividends far in excess of what they could really pay. The claims gained credence when some companies failed in the 1850s. As Stalson puts it (p. 226),

> The suspicion of all companies which these competitive assaults on individual companies engendered, however, unquestionably did every company more harm that it did any individual agent or company good.

It was about this same time (the early 1850s) that the states began to impose minimum capital requirements on life insurance companies and the first reserve-valuation laws were passed.

If the "lemon" theory of solvency regulation has much truth in it then it will be very difficult to assess regulation's benefits. One would have to estimate what the size of the market would have been in the absence of regulation. I doubt that this can be done, though it may be worthwhile to see whether differences in solvency regulations among the states result in any detectable differences in the size of their markets. But I don't think the interesting policy questions concern whether or not solvency regulations should be eliminated. Neither industry nor consumers appear to be pushing for their removal. Rather, I think the interesting policy questions concern whether the methods that have been adopted to achieve a given probability of insolvency are low-cost ways of doing so, and how one would go about deciding on an appropriate insolvency rate.

The first question, the relationship between paid-in capital (K), the premium "loading" (k), and the probability of insolvency or "ruin" (S), has been the subject of a great number of theoretical articles under the general title of "collective risk theory," and a much smaller number of empirical studies. (For a general review of these see Seal, *Stochastic Theory of Risk Business* [New York: Wiley, 1969], chapters 4 and 5, and K. Borch, *The Mathematical Theory of Insurance* [Lexington, Mass.: Lexington Books, 1974], parts III and IV.) The theoretical literature provides, under some simplifying assumptions, some fairly tractable closed-form analytical expressions for the function $s = g(k,K)$.

Some insight may be gained into the second question, the appropriate level for an insolvency rate, by exploring further the simple example I gave to illustrate the workings of the lemon principle. The rational regulator (a species at least as rare as the rational economic man) might reason as follows: For any given insolvency rate(s), I can use the technique illustrated in the figure to compute the "certainty equivalent" level of wealth for each policyholder. I can change s by requiring companies to change the "loading" they built into their rates or by changing the minimum capital requirement. What "loading" and minimum capital requirement should I require, if I want to maximize the sum of the policyholder's wealth certainty equivalents? The answer to this question could be found by solving the following maximization problem:

$$\max_{k,K} \sum_{i=1}^{n} \left[U(W_i - P)(1 - s) + U(W_i - P - C)(s)\right],$$

where

$s = g(k, K)$ and $P = f(k, K)$.

In solving this problem, the regulator could look explicitly at the tradeoff between increasing solvency and decreasing the premium. Note that the expected utility hypothesis is being used for a "normative," rational policy purpose, rather than for a "positive" purpose, such as to predict the way people actually behave.

Summary

I have suggested

- that solvency regulations in the life insurance industry deserve at least as much attention as those in the property and casualty lines,
- that Munch and Smallwood should look to Akerlof's "market for lemons" for a theoretical framework, rather than to the CAPM, and
- that the focus for policy research should be on the tradeoff between lower premium rates and lower insolvency rates, rather than on whether there should be any solvency regulations at all.

The views expressed in this comment are solely those of the author and do not necessarily represent those of the FTC or any of its other staff members.