Journal of Risk and Uncertainty, 6:187-203 (1993)

Insurance with Undiversifiable Risk: Contract Structure and Organizational Form of Insurance Firms

NEIL A. DOHERTY
Department of Insurance and Risk Management, The Wharton School, 3641 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104

GEORGES DIONNE*
Department of Economics and CRT, Université de Montréal, Canada

Abstract

Previous explanations of the contract choice and organizational form of insurance firms do not explain, by themselves, the recent proliferation of mutuals and new contract designs. We first present risk-bearing arguments to address these phenomena. We present two forms of insurance. The first is a conventional transfer of risk whereas the second decomposes risk between idiosyncratic and nonidiosyncratic. We show that the latter form leads to more active trade in insurance markets with correlated exposures. Moreover, the decomposed form dominates the simple transfer. These results qualify and extend the work of Borch (1962) and Marshall (1974). Market responses to the recent "liability insurance crisis" are compatible with these predictions.

Key words: aggregate risk, idiosyncratic risk, insurance contracts, mutualization principle, homemade mutualization, liability insurance crisis.

In recent years, there have been innovations in contract design and organization form that are not easily explained by the existing literature. In liability and earthquake insurance, there has been a proliferation of new firms such as mutuals, reciprocals, group captive insurance companies, and risk retention groups. The essential feature of all of these organizational forms is that they are owned by their policyholders.[1] For example, Danzon (1985) documents a dramatic increase in the market share of the mutuals in the medical malpractice insurance market after the mid-1970s.[2] A more dramatic example is the pollution insurance market, which all but disappeared in 1984–1985 only to show some signs of revival with the emergence of "mutual-like" pools organized by the major brokers (see Berkowitz, 1987). Similar innovations have also appeared in earthquake insurance. These organization structures share the common feature of combining the

*This research was conducted while G. Dionne was visiting the University of Pennsylvania under grants from Huebner Foundation and CRSH Canada. N. Doherty wishes to acknowledge with thanks support from the National Science Foundation under grant #88-09299. We wish to thank Patricia Danzon, Beverly Dahlby, Claude Fluet, James Garven, Scott Harrington, Howard Kunreuther, Clifford Smith Jr., George Theall, Kip Viscusi, and a referee for helpful comments.

equityholder and policyholder functions, thereby allocating residual claims on the insurance pool to the policyholders. Risk is pooled amongst those who are commonly exposed rather than transferred to external risk bearers. New policy forms also have been introduced by existing stock insurers. Under "claims made" liability policies, the policyholder is exposed to much of the risk of changing liability rules. This is similar in effect to mutualization.

The various markets in which these innovations have appeared share a common feature. In earthquake insurance, losses are correlated within any seismic zone, and risk will not be eliminated by pooling. Liability insurance suffers similar problems of nonindependence if the rules under which liability is determined and damages are assessed change over the life of the insurance contracts. These changes, introduced by legislation or judicial precedent, redefine the insurer's liability for all current policies.[3]

The seeds of an explanation for the innovations in contract design can be found in Borch's classic 1962 paper, "Equilibrium in Reinsurance Markets." Borch showed that the Pareto optimal risk sharing in an economy with risk-averse actors is one in which each shares in the aggregate wealth. If aggregate wealth is riskless, e.g., if there is a large number of actors and if the risk in individual endowments is uncorrelated, then individuals can fully insure. When aggregate wealth is risky, individuals may insure idiosyncratic risk but retain shares in aggregate wealth. This principle, sometimes called the *mutuality principle*, was later developed by Wilson (1968) in his Theory of Syndicates. However, it has largely been ignored in the extensive literature on optimal insurance contract design (e.g., Arrow, 1963; Raviv, 1979; see Gollier, 1992, for a survey). In that literature, the central problem is to derive the optimal contract between an insured and insurer in which price is specified as a function of the loss distribution of the policyholder, transaction costs, and a profit markup. Notable in this approach is the absence of a decision variable to allocate social risk.[4] An exception is an insightful paper written by Marshall over a decade ago. Marshall (1974) identified two principles under which insurance might function: the reserve, or transfer, principle and the mutualization principle. Under the reserve principle, risk is transferred to external risk bearers to hold in a reserve from which to discharge claims. With mutualization, policyholders jointly hold the residual claims on the pool. Total losses are shared among policyholders by some combination of prepaid premium and retroactive dividend. The reserve principle is efficient when, by the law of large numbers, the average loss is predictable with virtual certainty while the mutualization principle can be used in more general circumstances.

This article first examines alternative contracting arrangements for insuring personal risks when insurance firms are not able to eliminate risk by pooling. This circumstance may arise at least in two ways. First, aggregate wealth in the pool is risky. We will derive our analysis in this way. However, given that firms may exist for reasons such as transaction costs (Jensen and Meckling, 1976, etc.) insurers may face specific costs of bearing risk (Benston and Smith, 1976; Smith and Stutz, 1985; etc.). The second purpose is to rationalize the emergence of new contract and organization forms in the insurance industry. Two types of contracting are presented. In one arrangement the risk is simply transferred, in full or part, to an external risk bearer. This arrangement is naturally achieved by a stock insurer writing prepaid insurance policies. In the second arrangement, risk is decomposed into idiosyncratic and nonidiosyncratic risk, and separate allocations of these components are made between policyholders and external risk bearers. The alternative arrangements may be viewed in a different light. In the first form, the residual claims on the insurance pool are held by external risk bearers and, in the second, the residual claims are held, at least in part, by the policyholders. This second arrangement can be embodied in the choice of organizational form for the insurance firm (e.g., a mutual insurance company), in the type of contracts sold (e.g., in a participating policy sold by a stock firm), or in the personal portfolio choices made by individuals (what we later describe as *homemade mutualization*).

Basic assumptions and notation are presented in section 1. In section 2, efficient allocations of risk will be established and compared under the two forms of contracting transfer. Section 3 shows that the alternative methods of assembling the respective contracts may involve simple contract design or choice of organizational form. The comparison between the alternative forms of contracting assumes that the allocation of resources is unaffected by the choice of organization (i.e., the comparison relates only to allocation of risk and abstracts from the role of agency costs). Section 4 illustrates how the alternative contract designs have emerged in the marketplace. A conclusion highlights the main results concerning risk sharing in the presence of aggregate or social risk and discusses implications for markets other than insurance.

1. Definitions and assumptions

In this section, we set up a simple characterization of risk to show how the mutuality principle can be embodied in a common contract design. Consider a group of n individuals, each of whom is exposed to the chance of a random, but insurable, loss. The realized value of the loss of any individual i can be represented as the product of the aggregate loss suffered by the group, $L(=\Sigma L_i)$, and the share a_i of that individual in the group aggregate loss:

$$L_i \equiv a_i L. \tag{1}$$

The share a_i is a random variable. Ex post, the share of the individual will be zero if (s)he has no loss and can assume many values according to the size of any revealed loss. The properties of the aggregate loss, L, are important. With independence, L/n will converge on a known value with zero variance as n grows larger. The variance of L/n will not converge to zero if losses are heavily correlated.[5] If the aggregate loss is transferred to external investors, it will command a proportional risk premium.[6] The presence of a risk premium is important for our model and will be explained later.

The economy we envision contains risk-averse individuals, each having an endowment containing a riskless asset, risky but marketable assets, and a risky but nonmarketable asset. Individuals cannot issue claims on their nonmarketable assets, but they can purchase insurance. The insurable loss is assumed to be uncorrelated with marketable assets in the individual's portfolio.[7] The loss distributions of insureds are assumed to be identical. When we describe the *representative* insured, we are describing all insureds.

The economy also contains *investors* who are endowed with a portfolio of marketable assets. Investors may sell insurance, in which case they assume residual claims on the insurance pool. They may hold residual claims in an insurance firm selling directly to policyholders or the claims of a firm that reinsures the direct insurer. Alternative insurance arrangements can appear in which external investors play no role, e.g., a simple mutual insurance firm that is wholly owned by its policyholders.

Transaction costs are important in our model. We need to be able to explain why insurance exists. For example, they are necessary to sustain the demand for insurance (large transaction costs on nonmarketable assets). A goal of this article is to examine the effects of interdependence on the optimal form of insurance contracting. To keep this focus, we initially assume away agency costs,[8] together with administrative costs[9] and information asymmetries.[10] We will relax these assumptions as we continue into sections 2 and 3.

Undiversifiable risk is presented in the simplest manner. Aggregate loss L follows a two-point distribution with probabilities p and $(1-p)$:

$$L = < \begin{array}{ll} L^{\mathrm{u}} & (p) \\ L_{\mathrm{d}} & (1-p) \qquad \text{where } L^{\mathrm{u}} > L_{\mathrm{d}}. \end{array} \tag{2}$$

Individual risks also are simplified. While the number of individuals is finite, we assume that they are large in number and, as we would expect when large numbers are insured, deviations of average loss from expected loss become quite unimportant. For each of the n individuals, the probability of loss is q. We can represent the distribution of losses in this population as a process consisting of the selection of qn individuals from the population of n. Each person selected suffers a loss that is a proportion, $1/qn$, of the total loss. Given the total loss from equation (2), the average loss is L/n. As is the case under the law of large numbers, average loss is much less uncertain than individual loss. In fact, average loss is fixed. Individual risks of this sort relieve us of explicit attention to limiting processes as numbers grow without bound, and they cause no real loss of generality.

The assumptions imply that

$$L_i = \begin{array}{ll} 0 & ;(1-q) \\ L^{\mathrm{u}}/(qn) & ;pq \\ L_{\mathrm{d}}/(qn) & ;(1-p)q. \end{array} \tag{3}$$

From equation (3), we see that there is an implied correlation between individual losses, L_i. It is straightforward to show that $\mathrm{COVARIANCE}(a_iL; a_jL) = E(a_i)E(a_j)\mathrm{VAR}(L) \neq 0$. It will be convenient to denote the individual share of loss as $k = (1/qn)$.

The respective states of nature are defined in table 1. Table 1 identifies the states of nature defined by the conjunction of idiosyncratic loss (i.e., whether 0 or kL) and social loss (whether L^{u} or L_{d}). Each cell labels the state of nature and notes its probability.

Table 1. The respective states of nature

		Aggregate loss (probability)	
		L^u (p)	L_d $(1-p)$
Insured's loss (probability)	0 $(1-q)$	State 1 $p(1-q)$	State 2 $(1-p)(1-q)$
	kL (q)	State 3 pq	State 4 $(1-p)q$

2. A comparison of alternative insurance contracts

In the preceding section, risk is found in the two state variables a_i and L. Risk arising from the randomness of the a_i is referred to as *idiosyncratic risk*, and we refer to risk inherent in L as *dividend risk*.[11] This nomenclature recognizes that idiosyncratic risk is diversifiable within the insurance pool but dividend risk typically falls on the pool's owners. Two contracts are defined that may be distinguished by how they address idiosyncratic and dividend risk. Both contracts are routinely encountered in insurance markets. However, while the literature on insurance contract design has focused on one of these forms, we argue that the other is more consistent with Borch's mutualization principle and is preferred under conditions of dividend risk. The first contract is denoted as a simple transfer. The representative policyholder chooses some proportion, α, of his loss to insure:

Simple Risk Transfer

Insurance coverage is αL_i where $L_i = a_i L$.

The insurer writing such policies will not be able to eliminate risk through simple diversification within the insurance pool if L is random. It also may be noted that the contract offers only a single decision variable to address an allocation problem in which the state space is described by two state variables. The second contract follows Borch's mutualization principle and decomposes risk according to the state variables. In this contract, the parties may choose both the proportion of idiosyncratic risk, α, and the proportion of dividend risk, β, that is allocated between the representative policyholder and the external investors. This resembles many insurance arrangements in which the residual claim is held either by shareholders and/or by policyholders who receive, respectively, shareholder and policyholder dividends.

Decomposed Risk Transfer

Insurance coverage is $\alpha\{[L_i - (1/n)(L - L_d)] + [(\beta/n)(L - L_d)]\}$

The first square bracket contains only idiosyncratic risk which, in the limit, can be removed by the insurer through diversification. The second square bracket contains the dividend risk. To see this, consider the aggregate payout, AP, for an insurer holding a portfolio of n policies:

$$\begin{aligned} \mathrm{AP} &= \alpha\{[\Sigma L_i - (L - L_\mathrm{d})] + [\beta(L - L_\mathrm{d})]\} \\ &= \alpha\{[L_\mathrm{d}] + [\beta(L - L_\mathrm{d})]\}, \qquad \text{since } \Sigma L_i = L. \end{aligned} \tag{4}$$

If $\beta = 1$, external investors would be left with a portfolio having variance $\alpha^2\mathrm{VAR}(L)$. If $\beta = 0$, external investors would bear no risk; the residual claims would fall on policyholders. In effect, this would be a simple mutual insurance firm. Moreover, the simple transfer turns out to be a special case of the decomposed transfer. This is seen by setting $\beta = 1$ in which case the contracts are identical.[12]

2.1. *Contracting with decomposed and simple risk transfers*

We first examine decomposed risk transfer (DRT) contracts under the assumption that the distribution of aggregate loss is known by all parties. The DRT contract is an insurance policy packaged with a residual claim on the insurance pool. As we stress later, there are various methods of assembling such contracts that may involve the choice of organization form for the insurer, or the type of contract design chosen by an insurer of a given organization type. The insured pays a premium and receives some proportion α of his individual loss. In addition, the policyholder contracts to bear a proportion $(1-\beta)$ of the dividend risk. The dividend is paid after losses are realized. If insureds bear all dividend risk, $\beta = 0$, their total contributions (premiums minus dividends) would equal the aggregate loss in the pool ex post. For example, the insured might prepay a premium equal to his share of the expected aggregate loss, $E(L)$, and receive (pay) a retroactive policyholder dividend should losses be more (less) favorable:[13]

$$P = (\alpha/n)\{E(L) + (L - E(L))\} = (\alpha/n)L. \tag{5}$$

However, the contracting arrangements also permit division of the dividend risk between insured and insurer. For example, the DRT might be sold by a mutual firm. The allocation of the dividend risk could then be achieved by controlling the amount of reinsurance purchased by the mutual from an independent reinsurer.[14] The net cost of insurance now comprises an advanced payment equal to the expected loss plus a risk premium R. This risk premium arises from the transfer of a portion β of the undiversifiable risk to external investors; thus the advance premium is $E(L) + \beta R$.[15] The dividend to policyholders is now confined to that proportion of the residual claim retained by the policyholders, $(1-\beta)(L-E(L))$. The net cost of insurance is

$$\begin{aligned} P &= (\alpha/n)\{E(L) + \beta R + (1 - \beta)(L - E(L))\} \\ &= (\alpha/n)\{L - \beta(L - L_\mathrm{d}) + \beta[p(L^\mathrm{u} - L_\mathrm{d}) + R]\}. \end{aligned} \tag{6}$$

The presence of a risk premium is important for our model. Such a risk premium may be motivated in two ways. One explanation for a risk premium arises from the presence of aggregate or social risk. If L is correlated with the portfolio of all assets, it will command a risk premium in the capital market. Such undiversifiable risk is termed *systematic risk*.[16] A second explanation for a premium loading lies in the presence of firm-specific costs of risk bearing. For example, with convex tax functions, the expected tax liability of a firm is positively related to the variance of its pretax income. In the absence of a perfect market for tax arbitrage, the value of the firm will decline as the variance of its pretax income increases unless this risk is priced.[17] A similar story can be told in relation to the transaction costs of bankruptcy, agency costs, or information asymmetries existing between managers of a firm and providers of capital.[18] Either the systematic risk or the firm-specific cost of risk bearing can be evoked to explain the pricing of aggregate risk, though we use the former to derive our results. Since the risk premium reflects the market price of risk, the value of the insurer is unchanged when policies are sold at this price. Thus, assuming investors can restructure their portfolios at minimal cost, optimal risk sharing under the DRT contract can be identified by maximizing the expected utility of the representative insured:

$$\begin{aligned} \mathrm{EU} = {} & (1-q)pU[W-(\alpha/n)\{L^{\mathrm{u}}-\beta(L^{\mathrm{u}}-L_{\mathrm{d}})+\beta(p(L^{\mathrm{u}}-L_{\mathrm{d}})+R)\}] \\ & + (1-q)(1-p)U[W-(\alpha/n)\{L_{\mathrm{d}}+\beta(p(L^{\mathrm{u}}-L_{\mathrm{d}})+R)\}] \\ & + qpU[W-(\alpha/n)\{L^{\mathrm{u}}-\beta(L^{\mathrm{u}}-L_{\mathrm{d}})+\beta(p(L^{\mathrm{u}}-L_{\mathrm{d}})+R)\} \\ & \qquad -(1-\alpha)L^{\mathrm{u}}/(qn)] \\ & + q(1-p)U[W-(\alpha/n)\{L_{\mathrm{d}}+\beta(p(L^{\mathrm{u}}-L_{\mathrm{d}})+R)\}-(1-\alpha)L_{\mathrm{d}}/(qn)] \end{aligned} \tag{7}$$

where U is the insured's utility function and W is endowed wealth. The utility function is strictly concave and satisfies the von Neumann-Morgenstern axioms. The first-order conditions are

$$\begin{aligned} \mathrm{EU}_{\beta} = {} & p[(L^{\mathrm{u}}-L_{\mathrm{d}})(1-p)-R][(1-q)U_1'+qU_3'] \\ & -(1-p)[p(L^{\mathrm{u}}-L_{\mathrm{d}})+R][(1-q)U_2'+qU_4'] = 0, \end{aligned} \tag{8a}$$

$$\begin{aligned} \mathrm{EU}_{\alpha} = {} & p[(L^{\mathrm{u}}-L_{\mathrm{d}})(1-p)-R][(1-q)U_1'+qU_3'] \\ & -(1-p)[p(L^{\mathrm{u}}-L_{\mathrm{d}})+R]\,[(1-q)U_2'+qU_4'] \\ & +(1/\beta)(1-q)[L^{\mathrm{u}}pU_3'+L_{\mathrm{d}}(1-p)U_4'] \\ & -(1/\beta)(1-q)[L^{\mathrm{u}}pU_1'+L_{\mathrm{d}}(1-p)U_2'] = 0. \end{aligned} \tag{8b}$$

Substituting condition (8a) into (8b) yields

$$L^{\mathrm{u}}p[U_3'-U_1'] = L_{\mathrm{d}}(1-p)[U_2'-U_4']. \tag{9}$$

Since wealth in states 3 and 1 differs only by the individual's uninsured loss and wealth in states 4 and 2 also differs by the uninsured loss, the respective brackets in equation (9) cannot assume different signs. This follows from monotonicity of the utility function. The only solution is found when $U_3' = U_1'$ and $U_4' = U_2'$, which implies that policyholders fully insure their idiosyncratic risk, i.e., $\alpha = 1$. Thus individuals are indifferent between states 1 and 3 and between states 2 and 4, since their individual loss will be paid by this contract.

But this does not imply that all risk is removed. Policyholders still maintain an equity stake in the pool. Dividend risk borne by policyholders is removed only when $\beta = 1$.

We now determine the allocation of dividend risk.[19] Since full individual insurance is optimal, ($\alpha = 1$), then $U_3' = U_1'$ and $U_4' = U_2'$. Substituting into EU_β yields

$$p(1 - p)(L^u - L_d)[U_1' - U_2'] - R[pU_1' + (1 - p)U_2'] = 0. \tag{10}$$

Except in the case in which $R = 0$, satisfaction of this condition requires that $U_1' > U_2'$, which implies $W_1 < W_2$. Thus the individual would retain some dividend risk, $\beta < 1$. It will be noticed that the purchase of full direct insurance of idiosyncratic risk ($\alpha = 1$) is independent of the R (equation (9)).

2.2. *Discussion*

The conclusion of section 2.1 is consistent with Borch's optimal sharing rule and corresponds to the model used by Mace (1991) to analyze the affect of aggregate uncertainty on consumption. Thus, changes in the cost of external risk bearing will not affect the purchase of insurance against idiosyncratic risk. The individual insurance decision is separable from the dividend decision. In our model, the separability arises from the independence of a_i and L in equation (1).

Under the DRT, the insured chooses to fully insure idiosyncratic risk but bears part of the dividend risk. How much dividend risk is borne by the policyholder depends upon his/her degree of aversion to risk and on the cost of external risk bearing. In contrast, the SRT presents the insured directly with the external cost of risk bearing when seeking insurance protection. Since this cost entails a proportionate loading in the insurance premium, it induces less than full insurance protection. Thus, insureds participate in their idiosyncratic risk. This comparison suggests that when DRT contracts are traded, there will be a more active market for insurance of idiosyncratic risk than when SRT contracts only are available. The implication is that the DRT form of contract is predicted to appear in insurance markets in which the losses of different policyholders are correlated.

The results are illustrated in figure 1. Consider the DRT. The top part of the figure considers only the determination of α starting with an arbitrary value of β such as $\beta = 0$ (i.e., the policyholder bears all dividend risk). The axes distinguish only between the loss states 1–2 (in which the individual suffers no loss) and 3–4 (in which a loss is suffered). Position X shows the "no insurance" endowment conditional on L being realized at L_d; the coordinates are the wealth levels shown in equation (7) for states 2 and 4 with $\alpha = 0$. Similarly, Y shows the wealth endowment if L is realized at L^u; thus the coordinates are the wealth levels for states 1 and 3 in equation (7) with $\alpha = 0$. Since wealth can be transferred from state 1 to state 3 (and from 2 to 4) at an actuarially fair price, we show the usual fair-price lines together with full insurance solutions on the 45° line. Note that there are two solutions. With position A the aggregate loss is realized at L_d, and a high dividend is paid to policyholders. With position B the aggregate loss is L^u, and a low (or

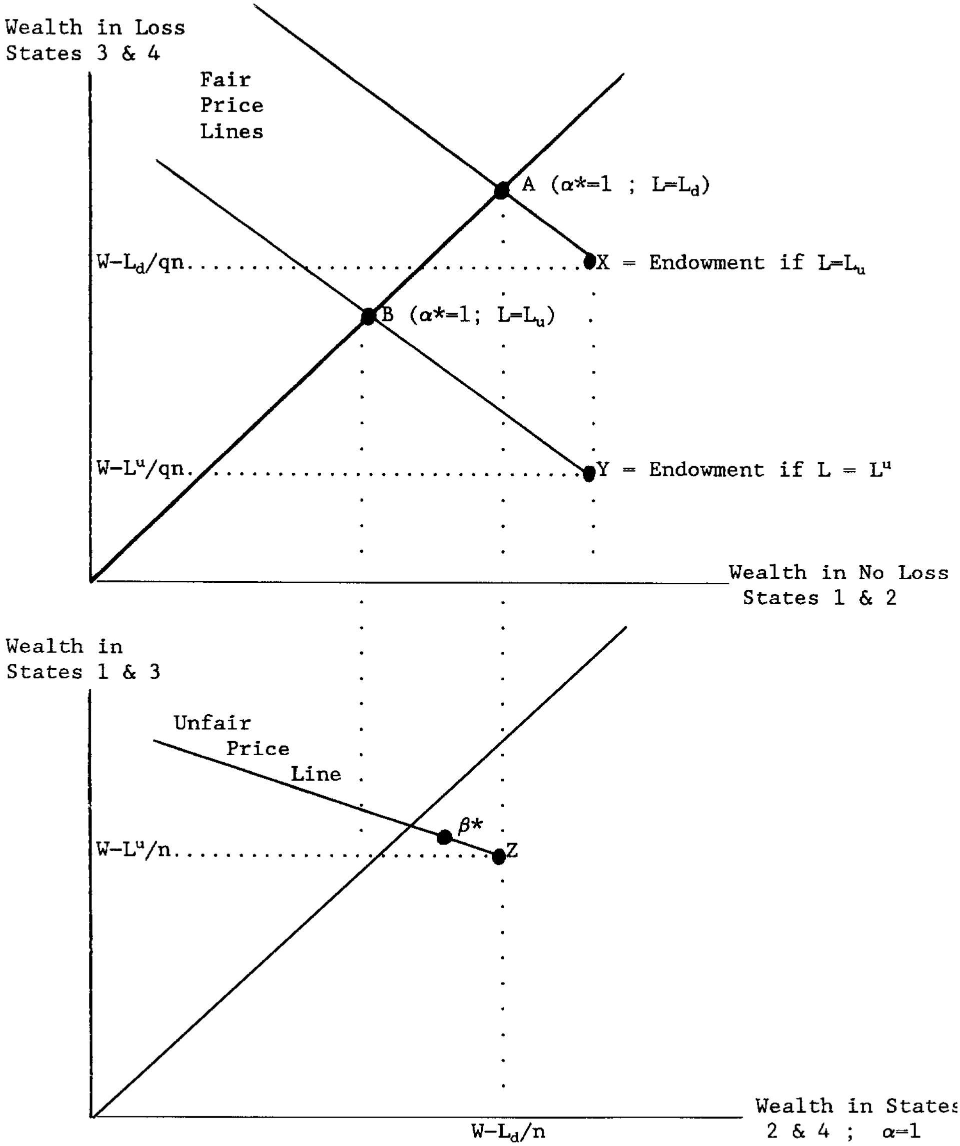

Figure 1. Separation of the insurance of idiosyncratic risk and the insurance of social risk.

no) dividend is paid to policyholders. The difference between A and B reflects dividend risk borne by the policyholder. The solution for α is independent of the dividend decision, which is shown in the bottom part of the figure.

The dividend decision is conditional on the solution, $\alpha^* = 1$, for idiosyncratic risk. Position Z shows the wealth coordinates for states 1-3 and 2-4 on the assumption that all dividend risk is borne by policyholders. Since an increase in β "insures" the dividend risk by transferring it from policyholders to external investors, this can be represented as a wealth transfer between states 1–3 and 2–4. Moreover, since the price of insuring dividend risk is unfair, $R > 0$, the price line will be flatter than the fair price line. The corresponding *efficient* dividend allocation, $\beta^* < 1$, is shown below the 45° line. If risk premium R is zero (e.g., if there is no firm-specific cost of bearing risk), the price line will be fair, and the optimal β will lie on the 45° line. Thus all dividend risk would be transferred to external investors. In this case, the DRT offers no advantage over the SRT contract. Marshall (1974, p. 484) has made a similar point for the case where aggregate loss is certain.

The dominance of the DRT contract is derived under a simple but restrictive characterization of loss distributions. The two variables used to describe idiosyncratic risk and aggregate risk were binomial. In combination, the effect is that the incidence of loss on individuals is described by the independently distributed a_i, but loss sizes are perfectly correlated. The convolutions of these variables for individuals reveal that individual wealth distributions are positively correlated but not perfectly so. These restrictions raise questions about the generality of our results. Our main result concerns the dominance of the DRT contract. It is fairly trivial to show that this result can be generalized. First, the simple risk transfer is simply a special case of the DRT contract.[20] This implies weak dominance. In fact, we can show that there is a mean-preserving spread between the two contracts. Second, it may be noted that our result illustrates the more general mutuality principle of Borch. What we have done is to focus on a particular contract design that satisfies this principle. Later we will argue that this contract design has indeed been adopted.

The second result concerns the separability of decisions on idiosyncratic risk and on aggregate risk. This result is not trivially generalized. The separability result depends on the independence of a_i and L, but not on the particular distributions. This result is a direct parallel of the "mutual fund separation," well known in finance, that permits separable decisions to control aggregate risk and diversifiable risk. The acid test, is of course, whether the contracts of the preferred form have emerged in markets with undiversifiable risk.

3. Contracting alternatives and homemade mutualization

The SRT and the DRT may be assembled in various ways. The assembly of these contacts can involve different forms of corporate organization or, alternatively, both forms can be assembled simply by a stock insurer through appropriate policy design. A simple policy sold by a stock insurer with a prepaid premium and no policyholder dividend would represent a SRT contract. In contrast, a policy sold by a mutual that did not reinsure would allocate all dividend risk to the policyholders. If the mutual purchased reinsurance from a second stock insurer, then part of the dividend risk would be transferred to the shareholders of the reinsurance firm. The extent of reinsurance protection thus controls the level of β in such a DRT. Reinsurance offers a secondary market for the dividend risk of mutual firms.

However, it is not necessary that a mutual type of organization sell the DRT contract. This was noted by Marshall (1974) in his comparison of the reserve and mutuality principles of insurance theory. He notes that stock firms might sell participating policies, which in his terminology are referred to as *mutual contracts*. We identify several possibilities for assembling the DRT that cut across forms of organization:

1. Various forms of organization naturally assemble the DRT. In addition to the mutual type of insurer, the reciprocal, or unincorporated mutual, is functionally similar, since here too the residual claims are held by the policyholders. Recent legislation responding to the liability insurance "crisis" has promoted the formation of similar structures. The Risk Retention Act (1986) permits the formation of risk retention groups that are subject to less severe regulation than other insurers. These firms issue inalienable stock to their policyholders and thereby resemble the mutual as discussed in section 5.
2. Stock insurers can assemble contracts that transfer an equity stake to the policyholder. One mechanism, most commonly used in life insurance, is the participating policy. The premium is subject to a retroactive adjustment that depends on the collective loss experience of the pool. This is simply a premium adjustment or a dividend. Though unusual in property liability insurance, such adjustments are sometimes encountered in workers' compensation insurance. A more widespread device that has similar, but not identical, effect is the *claims-made policy* used in liability insurance. This also is discussed in section 5 below.
3. The policyholder can, in principle, assemble the insurance and equity bundle on his own account. This is achieved by joint purchase of an SRT contract from a stock insurer and the traded equity of that insurer. This is labeled *homemade mutualization*. This idea is analogous to Modigliani and Miller's (1958) concept of *homemade leverage* and is valid only under comparable assumptions to those used by Modigliani and Miller. For example, taxes, contracting costs, and agency costs may have different effects on the investment decisions of the stock and mutual firms that cannot be undone by homemade mutualization.

Homemade mutualization is illustrated as follows.[21] The individual jointly purchases a policy from a stock firm with a coinsurance proportion α together with some proportion, π, of the equity of the insurer. Given limited liability and the limitations of the one-period model, it is natural to think in terms of the insurer paying a nonnegative dividend:

$$\text{Dividend} = \pi[L^{u} - L].$$

The purchase price of this equity share will reflect the risk premium, R:

$$\text{Price of equity share} = \pi[L^{u} - E(L) - R]$$

These equity cash flows can be combined with the policy cash flows to replicate exactly the payoff structure for the DRT contract defined in equation (7). This can be seen by the following general expression for payoff with homemade mutualization:

$$W - \alpha[(E(L) + R)/n - \pi(L^{u} - L) + \pi\{L^{u} - E(L) - R\}] - (1 - \alpha)L_i.$$

Rearrangement and substitution of $\pi = (1-\beta)/n$ yields the payoffs shown in equation (7). The solution is then identical to that shown in figure 1 with $\alpha^* = 1$ and $\beta^* = (1 - n\pi^*)$. The other properties of the solution for the DRT also are preserved. In particular, the *efficient* solution for the insurance contract and for the share of the insurer's equity are one-way separable, i.e., $\alpha^* = 1$ whatever the valueof π^*.

4. Emergence of DRT contracts

Here we illustrate how DRT contracts have emerged in one marketplace as that market has come to display increased aggregate risk. Priest (1987), Trebilcock (1987), Winter (1991), Doherty, Kleindorfer, and Kunreuther (1990), and others have argued that the recent liability insurance crisis is due largely to correlations between policy payoffs caused by a destabilization of liability rules. To add background, liability insurance typically has been offered on an "occurrence" basis. Such policies cover occurrences arising within the policy year. Often losses are discovered, and claims filed, years or even decades after the policy was written. After suits are filed, negotiations and litigation often proceed for months or years before final settlement is reached. These policies unite the policyholder and insurer in a long-term relationship. The delay exposes the firm to judicial, legislative, and economic changes that commonly influence the ultimate determination of awards. For example, unexpected inflation over the runoff period could commonly increase the ultimate payoff on all outstanding claims beyond the amount reserved by the insurer. Similarly, a new judicial precedent or legislation can expand the area of liability, ease the burden of proof for future plaintiffs, or cause a common upward shift in the value of awards on all outstanding claims. Such changes give rise to risk that cannot be diversified within the insurer's portfolio.[22]

The above analysis suggests alternative responses to the destabilization of liability rules. First, existing stock insurance firms wishing to survive in the affected lines would rewrite their insurance policies to shorten the period of the relationship and to ensure that policyholders participate in the residual value of the insurer. Second, new mutual forms of organization would emerge. Both responses have appeared, and we discuss each innovation in turn.

Many existing liability insurers have introduced new policy forms that provide cover on a *claims-made* basis. Though not universal, such forms are widespread in many liability lines. These policies cover only claims filed within the policy period, regardless of the timing of the occurrence that gave rise to the claim. Under the claims-made policy, the policyholder retains much of the undiversifiable risk arising from changing liability rules. Recalling the long settlement delays, the resolution of the claims reflects the liability rules in force at the time of resolution rather than the rules in force when the policy was written. The occurrence policy assigns the risk from such cumulative rule changes to the insurer. By shortening the period, the claims-made policy transfers the risk of changing liability rules on the long latency period claims to the insured (see Doherty, 1991). Our analysis favors forms of contracting that assign part of this risk to the policyholder.

The second predicted innovation is that new "mutual-like" firms would emerge where policyholders are able to monitor managerial performance at low cost. In fact, there has been a proliferation of such firms, most notably, the risk retention groups (RRGs). The RRGs are insurance pools that are established under the terms of the Liability Risk Retention Act of 1986. The Act facilitates the formation of pools of similar policyholders facing common liability insurance needs. The mutual nature of the RRG is ensured by requiring inalienable residual claims to be held by the policyholders. The restriction on alienability permits the capture of the risk-bearing advantages of the mutual type of contract.

Our analysis suggests that contract design or organization design can be used to embody the mutualization principle. While risk-bearing arguments suggest that these are close substitutes, other issues are involved in this choice. With costless contacting, the Coase theorem implies that organization form is irrelevant. With costly contracting, specific organization forms will have survival value where the combination of agency costs and production technology permit them to offer products at lower prices than other forms of organization (Fama and Jensen, 1983a, 1983b; Mayers and Smith, 1986, 1988). For example, the mutual form of organization eliminates the shareholder function and thereby removes opportunities for shareholders to expropriate policyholder wealth. However, the stock form of organization benefits from a set of capital market controls on managerial opportunism and incompetence. While risk-sharing arguments point to adoption of the mutualization principle in conditions such as those found in liability markets, whether this is accomplished by contractual or organizational devices rests largely on these agency issues.

5. Conclusion

Borch's subsequently labeled *mutualization principle* establishes that, in the presence of social risk, the Pareto optimal risk sharing arrangement is one in which idiosyncratic risk is fully insured but participants are left with a share of the social risk. This result suggests that an efficient insurance contract will decompose risk into diversifiable and nondiversifiable elements and will permit the parties to bargain on the sharing of each risk component. The DRT contract defined here was of this form, and we examined the optimal insurance decision for this type of contract. We showed that DRT contracts weakly dominate simple contracts that do not decompose risk, and that under some circumstances, the decisions on idiosyncratic risk and social risk are separable. The DRT contract may be assembled by deliberate contract design (a participating insurance policy), by organizational form (a mutual insurance firm), or by individual portfolio construction (homemade mutualization). In each case, the individual essentially bundles a simple insurance policy with an equity share on the insurance pool. Finally, we gave examples of how such innovations have recently appeared in markets characterized by increased social risk; in particular, such innovations have partly mitigated the effects of the liability insurance crisis.

Our results have been presented in the context of insurance markets. Thus the optimal DRT contract asserts that individuals will share in the aggregate cost of an earthquake or hurricane or in the aggregate of liability losses. However, as noted by Borch and subsequently embodied in the capital asset pricing model, similar conclusions can be reached about the sharing of risky aggregate societal wealth (GNP) even though much of this risk is not subject to insurance trade. Recent research by Mace (1991) and others have picked up on this theme. For one specification, Mace shows that changes in individual consumption are more closely related to aggregate risk than to idiosyncratic risk, since the latter is assumed to be fully insured. This result suggests independence between the two forms of risk. However, the results with another specification are less conclusive for the full insurance of idiosyncratic risk. Cochrane (1991) also obtained mixed results. One explanation is that full insurance is not the appropriate alternative in the presence of transaction costs and asymmetric information. A more appropriate test would be to compare the relative weights of both forms of risk on the variation in individual consumption. Another explanation is nonindependence between the two forms of risk.

Contract designs and trading strategies are available in noninsurance markets to share aggregate risk; for example, investors can trade in market indices or participate in mutual funds. Comparison of optimal trading strategies with optimal insurance, and comparison of the available instruments in capital and insurance markets, appears to offer some promise in helping us to understand the optimal allocation of social risk.

Notes

1. *Mutuals* are incorporated firms in which policyholders have rights similar to those of shareholders of joint stock firms; i.e., they receive dividends and have voting rights. *Reciprocals* are unincorporated mutuals. *Group captives* are typically stock firms, but stock ownership is confined to a limited number of policyholders (typically firms) who are insured by the captive. *Risk retention groups* are mutual-like forms set up under specific legislative provisions.
2. Danzon (1984) has also noted that the formation of mutuals and captives in the medical malpractice insurance market is consistent with the pricing of nonindependent risks under the risk transfer contracts. However, she does not develop this theme to show that mutuals and captives are optimal risk forms in such circumstances.
3. An insightful survey of issues in the particular troubled pollution insurance market is given in a report of the National Association of Insurance Commissioners (see Haayen, 1983). Of particular significance are the recent cases of mass torts and toxic torts (e.g., Dalcon Shield, asbestos, Bhopal, etc.), which challenge courts to create new, and perhaps radically innovative, liability rules. These innovations redefine cover, not only on current policies, but sometimes for policies issued years or decades earlier. Rabin (1987) discusses many of the legal implications.
4. These models conditioned the insurance payout solely on the individual policyholder's loss, whereas a participating policy conditions payout both on the individual loss and on the portfolio experience.
5. Strict independence is sufficient, but not necessary, for the law of large numbers (see Marshall, 1974), for more details.
6. Premium loadings to cover transaction costs are encountered in proportional and nonproportional forms. Risk premiums to compensate external investors for assuming risk are central to the contract comparison. In footnote 15 below, we show an example of such a risk premium using the capital asset pricing model.
7. While it is possible to generate similar results to those in section 2 without zero covariance between insurable and noninsurable assets (proof available on request), this assumption allows us to separate the insurance decision from other portfolio decisions of the insured (see Mayers and Smith, 1983). An

equivalent result may be derived for the class of risk-averse, state-independent utility functions following Doherty and Schlesinger (1983). See Eeckhoudt and Kimball (1992) for an extension of this analysis.

8. See Fama and Jensen (1983a, 1983b) and Mayers and Smith (1986, 1988).
9. See Arrow (1963) and Raviv (1979).
10. See, for example, Arrow (1963), Akerloff (1970), Marshall (1976), Dionne and Lasserre (1987), Smith and Stutzer (1990), and Dionne and Doherty (1991).
11. If L is correlated with the aggregate wealth, then *dividend risk* might be more conventionally labeled *social risk* or *aggregate risk* (see, for example, Allais, 1953; Borch, 1962; Dreze, 1971). However, we also consider the possibility that L is uncorrelated with aggregate wealth but, due to transaction costs, it is still costly for the insurer to bear this risk. The term *dividend risk* captures both cases and focuses on contractual choices for allocating this risk between policyholders and investors.
12. Another interpretation of equation (4) suggested by a referee is possible. Equation (4) refers to the insurer's aggregate portfolio loss, not to the payout on a single policy. Thus, one can think of the case where all contracts issued by the insurer have $\beta=0$ as a situation in which the insurer has an upper limit *reinsurance* contract with another insurance firm. Such contracts (known as stop loss contracts) are very common in the reinsurance market. However, with $\beta=0$, the contract the policyholder holds with the primary insurer would not be an upper limit contract; it would simply be a nonparticipating contract.
13. Some insurers cannot make negative dividends (assessments) under their terms of incorporation. Others, known as assessment mutuals, do not charge an advance premium but rely solely on a retroactive assessment to pay claims. If contracting is costly, a high advance premium provides a way of bonding policyholders to fulfill their contracts, i.e., not to default on payment. This is consistent with the comparative rarity of assessment mutuals.
14. The allocation of the dividend risk in the DRT contract corresponds to a form of reinsurance contract that elsewhere has been shown to be Pareto optimal; see Buhlmann and Jewell (1979) and Raviv (1979).
15. The allocation of the risk premium across policyholders is illustrated as follows. Suppose, for example, the risk premium arose because total losses, L, were correlated with the market portfolio. The risk premium demanded by investors, assuming the capital asset pricing model was appropriate, would be

$$\beta_R = \mathrm{COV}[\beta(L - L_d); R_m][E(R_m) - R_f]/\mathrm{VAR}\,(R_M),$$

where R_m is the return on the market portfolio and R_f is the risk-free rate. For each policyholder, the risk premium would be

$$\mathrm{COV}[(\alpha/n)\beta(L - L_d); R_m][E(R_m) - R_f]/\mathrm{VAR}(R_M) = (\alpha/n)\beta_R.$$

Notice that the risk premium paid by an individual increases with α. Similar examples of increasing risk premiums can be developed using other asset pricing models or using a firm-specific cost of risk bearing to motivate the risk premium.
16. The pricing of systematic risk is explained in asset pricing models such as the capital asset pricing model (CAPM) and the arbitrage pricing theory. It will be noted that the riskiness in L is a necessary, but not a sufficient, condition for the presence of a risk premium with these models. Recent evidence of nonzero underwriting betas is provided by Cummins and Harrington (1985). Their results, derived using insurers' quarterly underwriting data, do, however, reveal some intertemporal instability.
17. See Smith and Stultz (1985) and Scholes and Wolfson (1987). Alberts and Hite (1983) show how such imperfections (they specifically examine leverage related imperfections) affect product prices and quantities given competition in the product market.
18. See Greenwald and Stiglitz (1987, 1990).
19. A standard result in the contingent claims literature is that full insurance would be purchased in a complete market. However, our result is that only idiosyncratic risk is fully insured. The retention of dividend risk is incompatible with a complete market.
20. In introducing the SRT and DRT contracts, we showed that the DRT degenerated into the SRT when the value of β was constrained to unity. This constraint leaves the SRT with a single control variable, α, to span

the state space defined by the two state variables, a_i and L. It follows immediately that the SRT cannot be an optimal contract except in the special case that $R=0$. It is well known from elsewhere that, under a simple risk transfer, the optimal level of insurance is $\alpha^*=1$ if $R=0$ and $\alpha^*<1$ if $R>0$. These results can be confirmed easily by setting $\beta=1$ and maximizing equation (7) with respect to α.

21. Discussion of homemade mutualization requires a qualification of the assumptions. Separability of the demand for insurance from other portfolio decisions was achieved by assuming that the individual's insurable risk was independent of other items in his wealth portfolio. The assumption of zero covariance between marketable and nonmarketable assets effectively rules out the individual taking a position in his insurer's stock when the payoff on the policy and the payoff on the insurer's stock are correlated. However, the point here is that homemade mutualization is a joint purchase of the policy, and the insurer's equity has an identical payoff to the purchase of a policy from a mutual. Thus the assumption that the payoffs from buying a mutual policy are independent of background portfolio wealth is equivalent to assuming that the payoff to the homemade mutualization bundle is independent of the other marketable assets.
22. Examples of both judicial and legislative innovation that have had such effects are the "Jackson Township" case and the "Superfund" legislation. Both are claimed to have considerably expanded the liabilities of those involved in the manufacture or disposal of hazardous materials. See Doherty, Kleindorfer, and Kunreuther (1990).

References

Akerlof, George. (1970). "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84, 448-500.

Alberts, William W., and Gailen L. Hite. (1983). "The Modigliani-Miller Leverage Equation Considered in a Product Market Context," *Journal of Financial and Quantitative Analysis* 18, 425-437.

Allais, Maurice. (1953). "Generalisation des theories de l'equilibre economique general ei du rendement social au cas du risque," *Econometrie*, 81-110.

Arrow, Kenneth J. (1963). "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53, 941-973.

Benston, George J., and Clifford W. Smith, Jr. (1976). "A Transactions Cost Approach to the Theory of Financial Intermediation," *Journal of Finance* 31, 215-231.

Berkowitz, Joan B. (1987). "EIL: Insurance Coverage Isn't Enough," *National Underwriter* (March), 57-68.

Borch, K. (1962). "Equilibrium in a Reinsurance Market," *Econometrica* 30, 424-444.

Buhlmann, Hans, and William S. Jewell. (1979). "Optimal Risk Exchanges," *Astin Bulletin* 10, 242-262.

Cochrane, John H. (1991). "A Simple Test of Consumption Insurance," *Journal of Political Economy* 99, 957-976.

Cummins, J. David, and Scott Harrington. (1985). "Property-Liability Insurance Rate Regulation: Estimation of Underwriting Betas Using Quarterly Profit Data," *Journal of Risk and Insurance* LII, 16-43.

Danzon, Patricia M. (1984). "Tort Reform and the Role of Government in Private Insurance Markets," *Journal of Legal Studies* 13, 517-549.

Danzon, Patricia M. (1985). *Medical Malpractice: Theory, Evidence and Public Policy*. Cambridge, MA: Harvard University Press.

Dionne, Georges, and Neil Doherty. (1991). "Adverse Selection, Commitment and Renegotiation: Extensions to and Evidence from Insurance Markets," Working Paper #805, CRT, Université de Montréal.

Dionne, Georges, and Pierre Lasserre. (1987). "Dealing with Moral Hazard and Adverse Selection Simultaneously," working paper, Center for Research on Risk and Insurance, University of Pennsylvania, Philadelphia, PA.

Doherty, Neil A. (1991). "The Design of Insurance Contracts when Liability Rules are Unstable," *Journal of Risk and Insurance* 58, 227-246.

Doherty, Neil A., Paul Kleindorfer, and Howard Kunreuther. (1990). "Insurance Perspectives on an Integrated Hazardous Waste Management Strategy," *Geneva Papers on Risk and Insurance* 57, 407-427.

Doherty, Neil A., and Harris Schlesinger. (1983). "Optimal Insurance in Incomplete Markets," *Journal of Political Economy* 91, 1045-1054.

Dreze, Jacques H. (1971). "Market Allocation under Uncertainty," *European Economic Review* 2, 133-165.

Eeckhoudt, Louis and Miles Kimball. (1992). "Background Risk, Prudence and the Demand for Insurance." In G. Dionne (ed.), *Contributions to Insurance Economics*. Boston: Kluwer Academic Publishers.

Fama, Eugene F., and Michael C. Jensen. (1983a). "Separation of Ownership and Control," *Journal of Law and Economics* 26, 301-325.

Fama, Eugene F., and Michael C. Jensen. (1983b). "Agency Problems and Residual Claims," *Journal of Law and Economics* 26, 327-349.

Greenwald, B. C., and J. E. Stiglitz. (1987). "Financial Market Imperfections and Business Cycles," National Bureau of Economic Research Working Paper 2494.

Greenwald, B. C., and J. E. Stiglitz. (1990). "Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior," *American Economic Review* 80, 160-165.

Gollier, C. (1992). "Economic Theory of Risk Exchanges: A Review." In G. Dionne (ed.), *Contributions to Insurance Economics*. Boston: Kluwer Academic Press.

Haayen, Robert. (1983). *Report of the Advisory Committee to the NAIC Task Force on Profitability and Investment Income*. Northbrooke, IL: Allstate Insurance Company.

Jensen, Michael C., and William H. Meckling. (1976). "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics* 3, 305-360.

Mace, B. J. (1991). "Full Insurance in the Presence of Aggregate Uncertainty," *Journal of Political Economy* 99, 928-956.

Marshall, John M. (1974). "Insurance Theory: Reserves versus Mutuality," *Economic Inquiry* 12, 476-492.

Marshall, John M. (1976). "Moral Hazard," *American Economic Review* 66, 680-690.

Mayers, David, and Clifford W. Smith, Jr. (1983). "The Interdependence of Individual Portfolio Decisions and the Demand for Insurance," *Journal of Political Economy*, 304-311.

Mayers, David, and Clifford W. Smith, Jr. (1986). "Ownership Structure and Control: The Mutualization of Stock Life Insurance Companies," *Journal of Financial Economics* 16, 73-98.

Mayers, David, and Clifford W. Smith, Jr. (1988). "Ownership Structure Across Lines of Property Casualty Insurance," *Journal of Law and Economics* 31, 351-378.

Modigliani, Franco, and Merton H. Miller. (1958). "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review* 48, 261-297.

Priest, George L. (1987). "The Current Insurance Crisis and Modern Tort Law," *Yale Law Journal*, 1521-1590.

Rabin, Robert L. (1987). "Environmental Liability and the Tort System," *Houston Law Review* 24, 27-52.

Raviv, Artur. (1979). "The Design of an Optimal Insurance Policy," *American Economic Review* 69, 223-239.

Scholes, Myron S., and Mark A. Wolfson. (1987). "Taxes and Organization Theory," working paper, Stanford University, Stanford, CA.

Smith, B. D., and M. J. Stultzer. (1990). "Adverse Selection, Aggregate Uncertainty, and the Role for Mutual Insurance Contracts," *Journal of Business* 63, 493-510.

Smith, Clifford W., Jr., and Rene Stultz. (1985). "The Determinants of Firms' Hedging Policies," *Journal of Financial and Quantitative Analysis* 20, 391-406.

Trebilcock, Michael J. (1987). "The Social Insurance-Deterrence Dilemma of Modern North American Tort Law: A Canadian Perspective on the Liability Insurance Crisis," *San Diego Law Review*, 929-1002.

Wilson, R. (1968). "The Theory of Syndicates," *Econometrica* 36, 119-132.

Winter, Ralph A. (1988). "The Liability Crisis of 1984-86 and the Economics of Competitive Insurance Markets," working paper, Yale Law School, New Haven, CT.