CAHIER 9101

AN INTRODUCTION TO INSURANCE ECONOMICS

by

G. Dionne¹ and S.E. Harrington²

¹ Economics Department and Centre de recherche sur les transports, Université de Montréal.
² School of Business Administration, University of South Carolina.

January 1991


Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l’aide et le soutien à la recherche. Ce cahier a également été publié au Centre de recherche sur les transports (publication no 743).
ABSTRACT

This essay reviews the main developments of insurance economics subsequent to the pathbreaking work of Arrow and Borch. The eight sections include articles on (1) utility, risk, and risk aversion, (2) the demand for insurance, (3) insurance and resource allocation (in which we include Borch, 1962, and Arrow, 1965), (4) moral hazard, (5) adverse selection, (6) insurance market structure and organizational form, (7) insurance pricing, and (8) insurance regulation.

key words: Insurance economics, demand for insurance, moral hazard, adverse selection, insurance market structure, organizational form, insurance pricing, insurance regulation.

RÉSUMÉ


Mots clés: L’économie de l’assurance, la demande d’assurance, le risque moral, la sélection adverse, les structures de marché, les formes d’organisation, la tarification d’assurance, et la réglementation de l’assurance.

Although the prevalence of risk in economic activity has always been recognized (Green, 1984), deterministic models dominated economic explanations of observed phenomena for many years. As a result, the economics of insurance has a relatively short history. In early work that formally introduced risk and uncertainty in economic analysis (von Neumann and Morgenstern, 1947; Friedman and Savage, 1948; Altus, 1953; Arrow, 1953; Debreu, 1953), insurance was viewed either as a contingent good or was discussed in relation to gambling. Before 1960, economic literature was largely void of analyses of the nature of insurance markets or of the economic behavior of individual agents in these markets.

During the early 1960s, Kenneth Arrow and Karl Borch published several important articles (Arrow, 1963, 1965; Borch, 1960, 1961, 1962) that can be viewed as the beginning of modern economic analysis of insurance activity. Two of these papers are reprinted in this volume. Arrow was a leader in the development of insurance economics, and more generally, in the development of the economics of uncertainty, information, and communication. Arrow (1965) presented a framework of analysis that explains the role of different institutional arrangements for risk-shifting, such as insurance markets, stock markets, implicit contracts, cost-plus contracts, and futures markets. All of these institutions transfer risk to parties with comparative advantage in risk bearing. In the usual insurance example, risk averse individuals confronted with risk are willing to pay a fixed price to a less risk averse or more diversified insurer who offers to bear the risk at that price. Since both parties agree to the contract, they are both better off.

Risk is seldom completely shifted in any market. Arrow (1963) discussed three of the main reasons that risk shifting is limited: moral hazard, adverse selection, and transaction costs. Arrow (1965) emphasized the problem of moral hazard and suggested that coinsurance arrangements in insurance contracts can be explained by this information problem. Arrow (1963) showed in the absence of moral hazard that full insurance above a deductible is optimal when the premium contains a fixed-percentage loading. He also proved that risk aversion on the part of the insurer is another explanation for coinsurance. Both results were extended by Raviv (1979) and others.

1Borch (1990, Ch. 1) reviews brief discussions of insurance contained in the works of Adam Smith and Alfred Marshall, as well as the role of uncertainty in Austrian economics.

2References reprinted in this volume are highlighted. Arrow (1963) is reprinted in Diamond and Rothschild (1978) and Borch (1960, 1961) are reprinted in Borch (1990).

3In the insurance economics literature, coinsurance refers to a contract in which the insurer pays a fixed proportion of any claim amount.
Borch (1960, 1961, 1962) also made significant contributions to the theory of optimal insurance. He developed necessary and sufficient conditions for Pareto optimal exchange in risk pooling arrangements. He also showed, in a general framework, how risk aversion affects the optimal coverage (or optimal shares) of participants in the pool. Although his formal analysis was in terms of reinsurance contracts, it was shown by Moffet (1979) that the same result applies for contracts between policyholders and direct insurers. Borch's formulation of risk exchange influenced the development of principal-agent models (Ross, 1973; Holmstrom, 1979), and it has led to many other applications in the insurance literature. More generally, Borch made many contributions to the application of expected utility theory to insurance and influenced the development of portfolio theory and its applicability to the insurance industry. Finally, Borch's contributions established some important links between actuarial science and insurance economics (Loubergé, 1990).

**Outline of this Volume.** The remainder of this introductory essay reviews the main developments of insurance economics subsequent to the pathbreaking work of Arrow and Borch. In the process, the articles included in this volume are introduced. The remaining eight sections include articles on (1) utility, risk, and risk aversion, (2) the demand for insurance, (3) insurance and resource allocation (in which we include Borch, 1962, and Arrow, 1965), (4) moral hazard, (5) adverse selection, (6) insurance market structure and organizational form, (7) insurance pricing, and (8) insurance regulation.

The selection of articles was based on several criteria including the significance of the contribution, the representativeness of the work, and the desire to include empirical as well as theoretical articles. The selection process also considered whether the level of mathematics employed was likely to be accessible to most readers. In a few instances, we showed a slight preference for articles in books that are not as readily available as those published in journals.

For the most part, neither this introductory essay nor the remainder of the volume attempts to cover the wide variety of applications of insurance economics in the areas of health insurance, life insurance and annuities, social insurance, and in the law and economics literature. Instead, we review significant applications and include several articles dealing with property-liability insurance. This approach is at least partially due to our taste (and expertise). However, these articles and our introductory discussion help to illustrate issues, concepts, and methods that are applicable in many areas of insurance.

**Utility, Risk, and Risk Aversion**

*The Expected Utility Model.* Although the theory of decision making under uncertainty has frequently been criticized since its formal introduction by von Neumann and Morgenstern, it has been very useful in the study of optimal insurance decisions. Until recently, the linear expected utility model was the standard paradigm used to formally analyze economic behavior under uncertainty and to derive applications in many fields such as insurance. With objective probabilities, three basic axioms are necessary to obtain the von Neumann-Morgenstern theorem: weak order, independence, and continuity. Given these three axioms (and some other technical assumptions), insurance policy A will be chosen over policy B if and only if \( E_u(U > E_u(U) \) (where \( E_u(U \) is the linear expected utility associated with policy I). With subjective probabilities, additional axioms must be introduced in order to obtain a unique subjective probability measure over the set of states and a utility function that is unique up to a positive linear transformation: state-independent preferences and "reversal of order", which rules out moral hazard (Anscombe and Aumann, 1963; Karni, 1985; Drèze, 1961, 1987).

Linearity in probabilities is directly associated with the independence axiom (Macrina, 1987). This axiom has been challenged by many researchers, including Allais (1953a) who presented a now classic example that violates linearity in probabilities (and thus the independence axiom). Nonetheless, a large number of fundamental results in insurance economics have been derived from the linear expected utility model. In fact, very few contributions use non-linear models (see, however, Karni, 1990a), and the classical expected utility model remains the most useful approach for applications in insurance. We have chosen to reprint Machina's article for two main reasons. First, the classical linear expected utility model is presented with a different perspective than in other articles on the subject (e.g., Drèze, 1974; Shoemaker, 1982). Second, and more importantly, problems with the traditional model and some of the proposed responses are discussed in detail.

---

See Lemaire (1990) for a survey of these applications.

See Boyle (1990) for a survey of Borch's scholarly contributions.
Measures of Risk Aversion. The Arrow-Pratt measures of absolute and relative risk aversion (Arrow, 1965; Pratt, 1964) are commonly used in analyses of insurance decisions. They measure both the intensity of an individual’s preference to avoid risk and variation in this intensity as a function of wealth. Given a von Neumann-Morgenstern utility of wealth function, $U(W)$ with $U'(W) > 0$ and $U''(W) < 0$ for risk aversion, these measures of risk aversion are useful in calculating the certainty equivalent of a risky situation and the corresponding risk premium $\Pi^*$, which can be interpreted as the largest sum of money an insured with a given utility function is willing to pay above the expected outcome (actuarially fair premium) to avoid the risk. Moreover, an insured with utility function $U$ is said to be more risk averse than another insured with utility function $V$ if $\Pi^*_U \geq \Pi^*_V$ when both face the same risky situation and have identical nonrandom initial wealth.

Finally, the absolute measure of risk aversion corresponding to a given utility function $-U''(W)$ is said to be non-increasing in wealth, $W$, if in the same risky situation, $\Pi^*(W_i) \geq \Pi^*(W_j)$ for $W_i \leq W_j$. A necessary condition for decreasing absolute risk aversion is that $U''(W) > 0$.

Measures of Risk. Another important concept in the analysis of optimal insurance behavior is the measurement of risk. Let $X$ and $Y$ be two random variables with respective distribution functions $F_X$ and $F_Y$. $F_X$ is a mean preserving spread of $F_Y$ (Rothschild and Stiglitz, 1970) if $E(X) = E(Y)$ and $E(U) < E(U)$ where $E(U)$ is the linear expected utility associated with the random variable $U$). Many insurance contracts with actuarially fair premiums can be interpreted in terms of a mean preserving spread because they reduce the spread of the loss distribution without affecting the mean. For example, full insurance (i.e., a contract that pays the full amount of loss) produces a global decrease in risk since it implies the comparison of a risky situation with a non-risky one (Meyer and Ormiston, 1989).

In some cases, Rothschild and Stiglitz’s definition of increasing risk is too general to generate non-ambiguous comparative statics results (Meyer and Ormiston, 1985). When this is the case, a particular definition of an increase in risk can be defined by imposing restrictions on the distribution functions representing the initial and final random variables in order to compare the optimal values of decision variables for each distribution function. In a recent article, Alarie, Dionne, and Eckhoudt (1990) show how this methodology can be applied to the optimal choice of insurance coverage. Several types of increases in risk that represent particular cases of mean preserving spreads are analyzed including a strong increase in risk (Meyer and Ormiston, 1985), a "squeeze of the distribution" (Eckhoudt and Hansen, 1980), "tail dominance" (Eckhoudt and Hansen, 1984), and a relatively strong increase in risk (Black and Bulkley, 1989). Meyer and Ormiston (1989) generalized another definition of increasing risk: the "stretching of a density around a constant mean" (Sandmo, 1970). This approach, which they characterized as involving "deterministic transformations of random variables", also represents a particular type of mean preserving spread. It has been applied to many economic decision problems, such as optimal output choice under uncertainty (Sandmo, 1971; Leland, 1977), optimal saving under uncertainty (Sandmo, 1970), optimal portfolio choice (Meyer and Ormiston, 1989), and optimal insurance decisions (Alarie, Dionne, and Eckhoudt, 1990).

Demand for Insurance.  

Basic Models of Coinsurance and Deductible Choice. Mossin (1968) and Smith (1968) proposed a simple model of insurance demand in which a risk averse decision maker has a total wealth ($Y$) equal to $W - L$ where $W$ is nonstochastic wealth and $L$ is an insurable loss. To illustrate this model, first assume that the individual can buy coverage at $0 \leq \alpha \leq 1$ for a premium $\alpha P$ where $\alpha$ is the rate of insurance coverage (the coinsurance rate), $\lambda$ (0 $\leq 1$) is the premium loading factor, $E(L)$ is the expected loss, and $P = \lambda E(L)$. It can be shown that the optimal insurance coverage is such that $0 \leq \alpha^* \leq 1$ for $P \geq \bar{P}$ where $\bar{P} = \lambda E(L)$ solves

$$E[U(Y + \alpha^*(L - \lambda E(L))] = E[U(Y)]$$

In this section, we limit discussion to the case where insurance premiums are exogenously determined. The general case is considered in the next section.
and where \( U \) is a von Neumann-Morgenstern utility function \((U'() > 0, \ U''() < 0)\) and \( \text{EU(Y)} \) is the level of utility corresponding to no insurance. Hence, if the premium loading factor exceeds one but is less than \( \lambda' \), partial coverage \((0 < \alpha' < 1)\) is demanded.

When \( \lambda = 1 \), \( \alpha' \) is equal to one and the maximum premium that a risk averse individual is willing to pay over and above the actuarially fair value of full insurance is the Arrow-Pratt risk premium \((\Pi')\). This premium solves

\[
U(W - E(L) - \Pi') = E[U(Y)]
\]

As shown by Pratt (1964), a more risk averse individual with utility \( V = k(U) \), \( k' > 0 \), and \( k'' < 0 \) will have a risk premium \( \Pi'' \) greater than \( \Pi' \).

Another important result in Mossin (1968) is that insurance coverage is an inferior good if the insured has decreasing absolute risk aversion. Under this assumption, there are two opposite effects on the demand for insurance when the loading factor \( \lambda \) increases: a negative substitution effect and a positive wealth effect. Hoy and Robson (1981) proposed an explicit theoretical condition under which insurance is a Gilfen good for the class of constant relative risk aversion functions. More recently, Brijs, Dionne, and Eechoudt (1989) generalized the Hoy and Robson (1981) analysis and provided a necessary and sufficient condition for insurance not to be a Gilfen good. This condition bounds the variation of absolute risk aversion so that the wealth effect is always dominated by the substitution effect. Finally, Alarie, Dionne and Eechoudt (1990) present sufficient conditions to obtain the intuitive result that an insured will increase his demand for insurance when a mean preserving increase in risk is introduced in the initial loss distribution.

Another form of partial insurance is a policy with a deductible (Mossin, 1968; Gould, 1969; Pashigian, Schkade, and Menefee, 1966; Schlesinger, 1981). For the above model, consider a general indemnity function \( I(L) \) and premium \( P = \lambda I(L) dF(L) \) where \( \lambda > 1 \) is again a proportional loading factor. Then it can be shown under the constraint \( I(L) \geq 0 \) for all \( L \), that for every \( P \),

\[
I'(L) = \begin{cases} 
L - D^* & \text{if } L - D^* \geq 0 \\
0 & \text{if } L - D^* < 0
\end{cases}
\]

where \( D^* \) is the optimal deductible.\(^{11}\) Since an insured bears some risk with the optimal contract it is reasonable to expect that a more risk averse insured would prefer a policy with a smaller deductible and higher premium. This result was proved by Schlesinger (1981) and Karlin (1965). Moreover, under decreasing absolute risk aversion, \( dD'/dW > 0 \) (Mossin, 1968). Also, it is possible to infer the degree of risk aversion of insurance buyers by observing their choices of deductibles (Drezé, 1981). The above results are generated under the assumption that the contract is free of default risk. With insolvency risk the above results do not in general hold but some qualitative results can be obtained with stronger utility assumptions (Doherty and Schlesinger, 1990).

**Optimal Coverage with Random Wealth.** If \( W \) is an uninsurable random variable rather than fixed, the optimal level of coverage \( (\alpha^*) \) depends on the statistical relationship between \( W \) and \( L \). If, for example, the correlation coefficient is a sufficient measure of the relationship between \( W \) and \( L \), Doherty and Schlesinger (1983) have shown that the Mossin (1968) and Smith (1968) result on the optimal coinsurance rate with fixed \( W (\alpha') \) is qualitatively similar to the case in which \( W \) and \( L \) are independent. That is, \( \alpha^* = 1 \) when the premium is actuarially fair and \( \alpha'^* < 1 \) when \( \lambda > 1 \). Moreover, Eechoudt and Kimb (1990) showed that \( \alpha^* \neq \alpha^* \) when \( \lambda > 1 \). Specifically, they showed that \( \alpha^* > \alpha^* \) when the degree of absolute prudence \((U''(U)/U')\) is positive and nonincreasing in wealth. This result was proved for any pair of statistically independent risks. They also analyzed optimal deductibles and showed, under the same conditions, that \( 0 < D^* < D^* \) where \( D^* \) is the optimal deductible when \( W \) and \( L \) are independent random variables and \( D^* \) is the optimal deductible with fixed \( W \). Hence, with independent risks, more coverage is demanded than with fixed wealth under both coinsurance and deductible contracts.

It was mentioned above that a more risk averse individual with utility \( V \) is willing to pay a greater risk premium for full insurance than a less risk averse individual with utility \( U \) when \( W \) is not random. This result also holds when \( W \) and \( L \) are independent random variables. For example, Kihlstrom, Romer, and Williams (1981) showed that a more risk averse individual with utility \( V \) will be willing to pay a higher premium than an individual with utility \( U \) if the absolute risk aversion for either individual for realized levels of \( W \) is nonincreasing in wealth.

If \( W \) and \( L \) are negatively (positively) correlated, high losses are likely to accompany low (high) values of \( W \). Doherty and Schlesinger (1983) showed in the case of a two-state

\(^{11}\text{The next section considers the optimality of coinsurance and deductible contracts when the insurance premium is not exogenously specified.}\)
marginal distribution that $\alpha^* > 1$ ($< 1$) when actuarially fair insurance is available for $L$. They also analyzed non-actuarially fair insurance prices. More details and more general results are outlined in Schlesinger and Doherty (1985).\footnote{See also Doherty and Schlesinger (1983a), Schlenenburg (1986), Turnbull (1983), Eeckhoudt and Kimball (1990) and Levy-Garboua and Montmarquette (1990).}

**Insurance, Portfolio Choice, and Saving.** Mayers and Smith (1983) and Doherty (1984) analyzed the individual demand for insurance as a special case of general portfolio hedging strategy. They introduced nonmarketable assets (such as human capital) in a capital asset pricing model to simultaneously determine the demands for insurance contracts and other assets in the portfolio.\footnote{See Kahane and Kroll (1985) and Smith and Buser (1967) for extensions of these models.} Mayers and Smith (1983) proposed sufficient conditions for a separation theorem between insurance contracts and other portfolio decisions. However, their analysis suggests that portfolio and insurance decisions generally will be interdependent. Consequently, full insurance is not necessarily optimal even when insurance is available at actuarially fair prices. This result is similar to that obtained by Doherty and Schlesinger (1983).

Moffet (1975, 1977) and Dionne and Eeckhoudt (1984) provided joint analyses of the saving (consumption) and insurance decisions in a two-period model. Dionne and Eeckhoudt (1984), which generalized Moffet's results, showed that under decreasing temporal risk aversion deposits and insurance are pure substitutes in the Hicksian sense. Moreover, in their two-decision variable model, insurance is not necessarily an inferior good. They also presented two alternative conditions under which a separation theorem holds between insurance and savings:\footnote{See Drøze and Modigliani (1972) for another sufficient condition on utility to obtain separation between consumption, portfolio, and insurance decisions.} actuarially fair insurance premiums or constant temporal risk aversion. The conditions differ from those of Mayers and Smith (1983) in their portfolio model of insurance decisions without consumption. This difference can be explained by the fact that Mayers and Smith considered a menu of risky assets while Dionne and Eeckhoudt (1984) considered only a safe asset. The latter study, which used a more general utility function than Mayers and Smith, is actually more closely related to the consumption-portfolio model developed by Sandmo (1969).

More recently, Briys (1988) extended these studies by jointly analyzing insurance, consumption, and portfolio decisions in a framework similar to that defined by Merton (1971).

The individual's optimal insurance choice is explicitly derived for the class of isoelastic utility functions. Not surprisingly, the properties of optimal insurance coverage are much more difficult to characterize than in models where insurance is studied in isolation or in the presence of either consumption or portfolio choice alone.

**Self-Insurance and Self-Protection.** Returning to the case of a single random variable $L$, market insurance can be analyzed in relation to other risk-mitigation activities. Ehrlich and Becker (1972) introduced the concepts of self-insurance and self-protection. Self-insurance refers to actions ($y$) that reduce the size (severity) of losses (i.e., $L'(y) < 0$ with $L''(y) > 0$ and while self-protection refers to actions ($x$) that reduce the probability (frequency) of accidents ($p''(x) < 0$ with $p''(x) > 0$). Ehrlich and Becker gave conditions under which self-insurance and market insurance are substitutes and conditions under which self-protection and market insurance are complements. In both cases, self-protection and self-insurance activities were assumed to be observable by insurers.\footnote{See Winter (1990) for an analysis of self-protection and self-insurance under asymmetrical information.}

While Ehrlich and Becker (1972) focused on the interaction between market insurance and activities involving either self-insurance or self-protection, they did not study in detail interactions between self-insurance and self-protection with and without the existence of market insurance. Boyer and Dionne (1983, 1989) and Chang and Ehrlich (1985) presented propositions concerning the choices among all three activities. When full insurance is not available, risk aversion affects the optimal choice of self-insurance and self-protection. While it seems intuitive that increased risk aversion should induce a risk averse decision maker to choose a higher level of both activities, Dionne and Eeckhoudt (1985) showed in a model with two states of the world that this is not always the case: more risk avverse individuals may undertake less self-protection.\footnote{See Hiebert (1989) and Briys and Schlesinger (1990) for extensions of their analysis.}

**Corporate Demand for Insurance.** Portfolio decisions also have implications for the demand for insurance by corporations. When corporations are owned by shareholders who can reduce their investment risk at low cost through diversification of their own portfolios, risk aversion by owners is insufficient to generate demand for insurance. Specifically, if shareholders can costlessly eliminate the risk of corporate losses in their own portfolio through portfolio diversification, the purchase of insurance by corporations can only increase shareholder wealth if it increases expected net cash flows by an amount that exceeds any...
loading in insurance premiums. Mayer and Smith (1982) analyzed the corporate demand for insurance from the perspective of modern finance theory (also see Mian, 1982; Mayer and Smith, 1990, and MacMinn, 1990). They discussed how bankruptcy costs; risk aversion by managers, employees, customers, and suppliers; efficiencies in claims administration by insurers; and a number of other factors each can provide an incentive for the purchase of insurance even when shareholders can costlessly eliminate risk through portfolio diversification. In a later study, Mayer and Smith (1987) considered the possible ability of insurance to increase shareholder wealth by mitigating the underinvestment problem that was originally analyzed by Myers (1977).

State Dependent Utility. The previous analyses have implicitly assumed that all commodities subject to loss can be valued in relevant markets. Examples of such insurable commodities include buildings and automobiles. For these commodities, an accident primarily produces monetary losses and insurance contracts offer compensation to replace these in whole or in part. However, there are other commodities for which good market substitutes do not exist. Examples include goods health, the life of a child, and family heirlooms. For these “commodities”, an accident produces more than monetary losses; it also has a non-monetary component (such as “pain and suffering”). Non-monetary losses can be introduced in a two-state model (I for no-accident and II for an accident) by using state dependent utility functions (Cook and Graham, 1977; Karni, 1985). Without a monetary loss, an accident is assumed to reduce utility if $U(W) > U(W)$ for all $W$ (where $U(Y)$ is the utility in state $i$). With a monetary loss ($L > 0), U(W) - U(W - L)$ measures the disutility of the monetary loss and $U(W - L) - U(W - L)$ measures the disutility of the non-monetary loss.

Marginal utility of wealth also depends on the state of the world. Three cases usually are considered: (1) $U_Y = U_0$ for all $Y$; (2) $U_Y > U_0$ for all $Y$; and (3) $U_Y < U_0$ for all $Y$ where $U_Y$ denotes $dU/dY$. It can be shown that $\alpha > 1$ for a policy with an actuarially fair premium. Thus, the individual will buy more (less) insurance than under state independent preferences when the marginal utility of wealth is greater (less) in the accident state than in the no accident state for all $Y$. Karni (1985) showed how an increase in risk aversion affects optimal insurance coverage when preferences are state-dependent, but the extension of measures of risk aversion to this case is not straightforward.

---

17This statement also holds if insurable risk has an undiversifiable (i.e., market) component, since insurers have no comparative advantage in bearing market risk (see Mian, 1982).

18This paper became a chapter in Debreu (1959).

19In the Arrow-Debreu world each agent has incomplete information about states but all agents share the same information (Radner, 1968). The latter implicit assumption rules out moral hazard and adverse selection problems.

providing insurance does not depend on the amount of coverage. Coinsurance was explained either by insurer risk aversion or convexity of insurer costs. Conditions for an optimal contract with an upper limit of coverage also were presented. All these results were obtained under the constraint that coverage be nonnegative.  

Kihlstrom and Roth (1982) studied the nature of negotiated insurance contracts in a non-competitive context in which there is bargaining over the amount and price of coverage. They showed that a risk neutral insurer obtains a higher expected income when bargaining against a more risk averse insured and that the competitive equilibrium allocation is not affected by the insured's risk aversion. Many of their results are represented in an Edgeworth Box diagram.

Moral Hazard

The concept of moral hazard was introduced in the economics literature by Arrow (1963), Drèze (1961), and Pauly (1968) (see also Kihlstrom and Pauly, 1971, and Spence and Zechauer, 1971). Two types of moral hazard have been defined according to the timing of an individual's actions in relation to the determination of the state of nature. They can be called ex ante and ex post moral hazard. In the first case the action is taken before the realization of the state of nature while in the second case the action is taken after.

Ex Ante Moral Hazard. Pauly (1974), Marshall (1976), and Shavell (1979) considered the case in which the occurrence of an accident (or the output of the consumption good) can be observed by the insurer and where neither the insured's actions nor the states of nature are observed. Under this structure of asymmetric information, the provision of insurance reduces (in general) the incentive to take care compared to the case of full information. Thus, there is a trade-off between risk sharing and incentives for care.

Shavell (1979) used a simple two-state model where the individual faces either a known positive loss or no loss with probabilities that depend on effort (care) to show that partial insurance coverage is optimal in the presence of moral hazard. He emphasized that the cost of care has a major impact on the optimal solution. Another important result was that moral hazard alone cannot eliminate gains to trade in insurance markets (i.e., it reduces but does not eliminate the benefits of insurance). These results were obtained assuming that the insurer has no information on an individual's level of care. In the second part of the paper, Shavell showed that moral hazard problems are reduced (but not eliminated) when actions are partially observable (also see Holmstrom, 1979).

Shavell's two-state model did not permit a detailed characterization of insurance contracts. More than two states are necessary to derive conditions under which deductibles, coinsurance, and coverage limits are optimal under moral hazard (see Holmstrom, 1979, and Winter, 1990, for detailed analysis).

Moral hazard in insurance also can be analyzed within a general principal-agent framework (Ross, 1973; Holmstrom, 1979; Grossman and Hart, 1983). However, certain conditions must be imposed to generate predictions. First, the action of the agent cannot affect the support of the distribution of outcomes, a condition naturally met in the two-state model (Shavell, 1979). The other two conditions concern the use of a first-order condition to replace the incentive compatibility constraint. The first-order approach is valid if it identifies the global optimal solution. Mirrlees (1975) and Rogerson (1985) proposed two sufficient conditions for the first-order approach to be valid when corner solutions are ruled out: (1) the distribution function must be a convex function of effort and (2) the likelihood ratio has to be monotone. If the distribution function satisfies the above conditions, optimal insurance coverage will be decreasing in the size of loss since large losses signal low effort levels to a bayesian principal. Jewitt (1988) recently questioned the intuitive economic justification of these two conditions and showed that they can be violated by reasonable examples. Specifically, he showed that most of the distributions commonly used in statistics are not convex. He then supplied an alternative set of conditions including restrictions on the agent's utility function to validate the first-order approach (see Winter, 1990, and Arnott, 1990, for further discussion).

Grossman and Hart (1983) proposed a method to replace the first-order approach. They also showed that the two conditions proposed by Mirrlees and Rogerson are sufficient to obtain monotonicity of the optimal incentive scheme. They analyzed the principal problem

---

21See Gollier (1987) for an extensive analysis of this constraint and Gollier (1990) for a recent review of optimal insurance contracting.

22The ex ante actions can affect event probabilities, event severity, or both (see Winter, 1990, for more details).
without using the first-order approach and consequently did not need any restriction on the agent's utility function. As Grossman and Hart noted, many of their results were limited to a risk-neutral principal. This restriction is reasonable for many insurance problems.\textsuperscript{24}

Long term contracts between principals and agents can increase welfare in the presence of moral hazard (Rogerson, 1985a; Radner, 1981; Rubinstein and Yaari, 1983; Boyer and Dionne, 1989a). In multiperiod insurance models, an individual's past experience eventually gives a good approximation of care. Hence insurers use the individual's past experience to determine premiums and to increase incentives for exercising care.

Moral hazard may alter the nature of competitive equilibrium by, for example, introducing nonconvexities in indifference curves. A competitive equilibrium may not exist, and when it does, insurance markets for some risks may fail to exist. More importantly, neither the first nor second theorems of welfare economics hold under moral hazard. Since market prices will not reflect social opportunity costs, theory suggests that governmental intervention in some insurance markets possibly could improve welfare if government has superior information (Arnot and Stiglitz, 1990; Arnot 1990).

Moral hazard also can affect standard analyses of government responses to externalities. An important example involves liability rules and compulsory insurance.\textsuperscript{25} With strict liability and risk averse victims and insurers, Shavell (1982) showed with perfect information that both first-party and liability insurance produce an efficient allocation of risk between parties in a model of unilateral accidents (with pecuniary losses only). When insurers cannot observe defendants care, moral hazard results in a trade-off between care and risk sharing (as in the case of first-party coverage). Shavell (1982) noted that if the government has no better information than insurers, its intervention in liability insurance does not improve welfare. This conclusion assumed that defendants were not judgement proof (i.e., they had sufficient assets to fully satisfy a judgement). Otherwise, their incentives to purchase liability insurance are reduced (Keeton and Kwerel, 1984; Shavell, 1986). Under strict liability, Shavell (1986) showed that if insurers cannot observe care, insureds buy partial insurance and the level of care is not optimal. He also showed that making liability insurance compulsory under these conditions need not restore efficient incentives. In fact, compulsory

\textsuperscript{24}See Dye (1986) and Mookerjee and Png (1989) for recent applications of Grossman and Hart's model.

\textsuperscript{25}See Danzon and Harrington (1990) for a survey on the demand and supply of liability insurance.

insurance could reduce care, and it is even possible that prohibiting insurance coverage could improve the level of care.

\textbf{Ex Post Moral Hazard}. The second type of moral hazard was first suggested by Spence and Zeckhauser (1971) who showed that an optimal contract between a principal and agent depends on the principal's ability to monitor the state of nature, the ex ante actions taken by the agent, and the nature of the accident. The previous discussion of ex ante moral hazard assumed that the principal knew the nature of the accident. Marshel (1976), Dionne (1984), and Townsend (1979) investigated the case in which the nature of an accident is not perfectly observable by the principal. Townsend (1979) considered the case in which the nature of the accident is known by the agent and verification is costly to the principal. One interpretation of such costly verification is auditing.

Mookerjee and Png (1989) extended the Grossman and Hart (1983) model to consider optimal contracts in the presence of both ex ante and ex post moral hazard. In their model, the agent takes an unobservable action that affects accident probabilities and then reports his realized accident to the principal. The principal may audit the report at a cost. Their main result is that random audits reduce expected auditing costs without distorting the incentives of the agent provided that wealth of the agent is strictly positive in all states of the world. Their results apply when falsification is costless and verification is costly. Lackey and Weinberg (1989) showed that partial insurance can be optimal if the nature of an accident can be falsified by the agent, but only at a cost.\textsuperscript{26}

\textbf{ADVERSE SELECTION}

Adverse selection occurs in insurance markets when information is asymmetric, i.e., when the insurer cannot observe an individual's risk at the time policies are issued and the individual has superior information about his or her risk. Akerlof (1970) proposed that if insurers have imperfect information about differences in risk for prospective insureds, then some insurance markets may fail to exist and others may be inefficient. Studies have

\textsuperscript{26}See Dionne and St-Michel (1988) for an empirical measure of the second type of moral hazard in the workers' compensation market.
analysed the ability of partial insurance coverage, experience rating, and risk categorization to reduce the negative effects of adverse selection.27

Partial Insurance and Sorting. Partial insurance coverage can result from two types of insurance pricing: "price only" policies (Pauly, 1974) and "price-quantity" policies (Rothschild and Stiglitz, 1976; Stiglitz, 1977). In the first case, insurers charge a uniform premium rate per unit of coverage to all applicants. Pauly's model ruled out price-quantity competition by assuming that insurers could not observe the total amount of coverage purchased by a client. In the second case, insurers offer a menu of policies with different prices and quantities so that different risks choose different insurance policies. These pricing strategies have been studied for single vs. multi-period contracts, for competition vs. monopoly, and, when assuming competition, for several different equilibrium concepts.28

In a single period model with competition, Rothschild and Stiglitz (1976) first showed that a pooling equilibrium cannot exist if a Nash definition of equilibrium is adopted (i.e., if each firm assumes that competitors' contract offers are independent of its own offer). Conditions under which "separating" contracts reveal information about insured risk were then studied by the authors. A major result is that when firms offer a menu of policies with different prices and quantities, policyholders may be induced to but do not necessarily reveal hidden information.29 They showed that a separating Nash equilibrium can exist in which high risk and low risk buyers purchase separate contracts. This separating equilibrium is characterized by zero profits for each contract, by partial insurance coverage for low risk buyers, and by full insurance for the high risk buyers. However, when there exist relatively few high risk persons in the market, they showed that neither a separating nor a pooling equilibrium exist.

Other equilibrium concepts that eliminate the non-existence problem have been proposed. Wilson (1977), Miyazaki (1977), and Spence (1978) (WMS) considered the case

27We only consider models in which uninformed agents move first (screening): uninformed insurers offer contracts and consumers choose contracts given their accident probability. Stiglitz and Weiss (1984) analyzed differences between screening and signalling models.

28See Cooper and Hayes (1987), Crocker and Snow (1985), and Cresta (1984) for an introduction to these models and Dionne and Doherty (1990) for a survey on adverse selection in insurance contracts.

29A similar analysis was provided by Stiglitz (1977) for the monopoly case. In his model there is always a separating equilibrium and the monopolist extracts all surplus subject to self-selection constraints.

in which firms anticipate that other insurers' policies that become unprofitable as a result of new offerings will be withdrawn.30 A WMS equilibrium is a pair of contracts in which profits on low risk contracts offset losses on high risk contracts. A WMS equilibrium exists regardless of the number of high risk persons in the market. If a Nash equilibrium exists, it coincides with the WMS equilibrium.31 Finally, a WMS equilibrium is always second best efficient.

Dahiby (1983) provided some empirical evidence of adverse selection in the Canadian automobile insurance market. He suggested that his empirical results were consistent with the WMS model with cross-subsidization between individuals in each class of risk. However, Riley (1983) argued that Dahiby's results were also consistent with Wilson's (1977) anticipatory equilibrium and Riley's (1979) reactive equilibrium. Cross-subsidization is not feasible in either of these models.

Experience Rating. Experience rating can be viewed as either a substitute or a complement to both risk categorization and sorting contracts with self-selection constraints when adverse selection is present.32 One polar case is when infinite length contracts yield the same solution as with full information. In this case, ex ante risk categorization is useless.

The other polar case is when costless risk categorization permits full observation of an individual risk so that information on past experience is irrelevant. While experience rating, risk categorization, and sorting contracts are used simultaneously in most markets, economic analysis to date has considered the three mechanisms independently (see Dionne and Doherty, 1990, for a more detailed review).

30The anticipatory concept of equilibrium was introduced by Wilson (1977) . Miyazaki (1977) (for the labor market) and Spence (1978) (for the insurance market) extended Wilson's model to the case in which each firm could break even by offering a portfolio of contracts. Riley (1979) and Grossman (1979) proposed other non-Nash equilibrium concepts. (See Crocker and Snow (1985) for a review of alternative equilibrium concepts).

31Each of these models either explicitly or implicitly assumed that insurers could enforce the requirement that their customers would buy coverage from only one insurer. Hellwig (1988) considered a model with endogenous sharing of information about customers' purchases and obtained an equilibrium with a reactive element that is similar to Wilson's (1977) anticipatory equilibrium.

32See Dahiby (1990), Dionne and Lasserre (1987), and Dionne and Vanasse (1988) for analyses of experience rating when moral hazard and adverse selection are present simultaneously.
Dionne (1983), Dionne and Lasserre (1985), and Cooper and Hayes (1987) extended Stiglitz's monopoly model (1977) to multi-period contracts. Dionne (1983) considered infinite length contracts without discounting while Cooper and Hayes (1987) mainly dealt with a finite horizon model (without discounting). While findings in both cases suggested that experience rating induced sorting or risk disclosure, the analyses differ in many respects. In Dionne (1983), a simple statistical review strategy is proposed along with risk announcement in the first period. The insurer offers a buyer full coverage at the full information price unless the observed average loss is greater than the true expected loss plus a statistical margin of error. Otherwise, full coverage is offered at a premium that includes a penalty. Both elements — announcement of risk and penalties — are necessary to obtain the same solution as with full information. They have the same role as the self selection constraint and the premium adjustment mechanism of Cooper and Hayes (1987). In their model, the premium adjustment mechanism served to relax the self-selection constraints and to increase the monopolist's profits. Finally, in both articles the monopolist commits to the terms of the contract.  

Cooper and Hayes (1987) also extended the Rothschild and Stiglitz (1976) model to two periods assuming that a Nash separating equilibrium exists. When consumers were assumed to be bound to a two-period contract, they obtained the same result as for the monopoly case. When the assumption that consumers sign a binding two-period contract was relaxed, they showed that competition in the second period limited but did not eliminate the use of experience rating. In both cases, the insurer was assumed to be committed to its experience rating contract.

Nilssen (1990) analyzed experience rating contracts without commitment by insurers in a competitive market. His results differed from those of Cooper and Hayes and were quite similar to those of Kunreuther and Pauly (1985), who assumed that insurers sell price-only policies (Pauly, 1974) rather than price-quantity policies. Another important assumption in Kunreuther and Pauly's model was myopic behavior by insureds, whereas firms could have foresight. With foresight, firms suffer losses in early periods, and make profits in later periods, whereas in the Cooper-Hayes (1987) model, they make profits in the initial period and losses in subsequent periods. D'Arcy and Doherty (1990) provided some empirical evidence that is consistent with Kunreuther and Pauly's model.

Risk Categorization. In most types of insurance, insurers classify risks using many variables. In auto insurance, for example, evidence indicates that driver age and sex are significantly related to accident probabilities (Dionne and Vanasse, 1988). In particular, evidence suggests that young male drivers (less than age 25) have much higher accident probabilities than the average driver. Since age and sex can be observed at very low cost, competition will force insurers to charge higher premiums to young males. Categorization using particular variables is prohibited in many markets, and the efficiency of categorization is an important policy issue.

Is statistical classification efficient in the presence of asymmetric information and adverse selection? Crocker and Snow (1985, 1986; also see Hcy, 1982, and Rea, 1987, 1990) showed that costless imperfect categorization always enhances efficiency when efficiency is defined as in Harris and Townsend (1981): second-best efficiency given the self-selection constraints imposed by asymmetric information. However, if classification is costly, the efficiency implications were ambiguous. Crocker and Snow (1986) also considered the existence of a balanced-budget tax-subsidy policy that provides private incentives to use risk categorization. With appropriate taxes, they showed that no agent would lose from classification. In their 1986 article, the results were shown using a WMS equilibrium, but a tax system also may sustain an efficient allocation with a Nash equilibrium. Their results can also be applied to a Wilson (1977) anticipatory equilibrium, or to a Riley (1979) reactive equilibrium (see Crocker and Snow, 1985). These results suggest that prohibiting statistical discrimination will impose efficiency losses in insurance markets when classification is virtually costless (e.g., age and sex classification in auto insurance).

Market Structure and Organizational Form

The seminal study by Joskow (1973) on market structure, conduct, and performance in the U.S. property-liability insurance industry considered market concentration and barriers to entry, estimated returns to scale, analyzed direct writer (exclusive agency/salaried employee) and independent agency (multiple insurer representation) distribution systems, and discussed possible effects of rate regulation on prices and availability of coverage. While written when rate regulation was predominant and when rating bureaus had a greater impact

---

23See Hosios and Peters (1989) for an analysis of contracts without any commitment by a monopolist in a finite-horizon environment.

24We limit our discussion to exogenous categorization of risks. See Bond and Crocker (1990) for an analysis of endogenous categorization of risks.
on the market than later in the 1970s and in the 1980s, this study nonetheless provided a basis for later work on a variety of subjects.\textsuperscript{35}

Concentration, Ease of Entry, and Consumer Search. Joskow concluded that market concentration levels were low, especially for the national market, and that significant entry barriers did not exist. He estimated simple models of insurer operating expense ratios and concluded that the industry was characterized by constant returns to scale. He did find, however, that expense ratios were much lower for direct writers than for independent agency insurers. Cummins and VanDerhei (1979) estimated more elaborate models than those employed by Joskow using pooled cross-section and time-series data. Their results again indicated significantly lower expense ratios for direct writers, but they suggested increasing returns to scale throughout the range of output.\textsuperscript{36}

While the results of other studies that have estimated cost functions with cross-sectional accounting data also suggest increasing returns to scale (e.g., Doherty, 1981; Johnson, Flanigan, and Weisbart, 1981), the use of accounting data to infer returns to scale is problematic. Among other limitations, available data on insurance company operating expenses aggregate capital (e.g., product and market development) expenditures and current costs. Firm output also cannot be measured accurately.\textsuperscript{37} Appel, Wonnall, and Butler (1985) analyzed changes in the size distribution of insurers over time. Their results were inconsistent with increasing returns for small insurers and thus more in line with evidence on entry and levels of concentration.

Joskow argued that differences in operating costs between direct writers and independent agency insurers could not be explained by differences in service.\textsuperscript{38} In order to explain why direct writers had not grown more rapidly, he suggested that prior approval rate regulation had discouraged price cuts by direct writers, that difficulty in raising capital and obtaining consumer recognition slowed their expansion, and that it would be costly for independent agency insurers to become direct writers. As a result, he concluded that direct writers behaved as oligopolists subject to short-run capacity constraints and that constrained profit maximization involved selection of risks with lower than average expected claim costs.

Smallwood (1975) also suggested barriers to insurers switching to direct writer distribution. He argued that independent agency insurers were more vulnerable to adverse selection, and he developed a formal model of insurer risk selection (which did not consider asymmetric information). However, in contrast to Joskow’s analysis, Pauly, Kleindorfer, and Kunreuther (1986) argued that significant barriers to raising capital for growth were highly unlikely. Instead, they suggested that direct writers and independent agency insurers produced different levels and types of services.

Joskow also conjectured that costly consumer search for low prices impeded direct writer growth.\textsuperscript{39} Joskow and others (e.g., Kunreuther, Kleindorfer, and Pauly, 1983) have suggested that search for low prices is costly because of differences among insurers in risk selection criteria and because information provided by friends and neighbours that have different risk characteristics may convey little information. In an empirical analysis, Dahly and West (1986) concluded that premium dispersion in Canadian auto insurance was consistent with a model of costly consumer search. This conclusion was contingent on their argument that risk classification could not account for premium variation. Berger, Kleindorfer, and Kunreuther (1989) modeled word of mouth transmission of price information in auto insurance in conjunction with consumer switch costs.

Returns to Scale and Underwriting Risk. The previously discussed studies of returns to scale and entry conditions focused primarily on insurer underwriting (risk selection).

\textsuperscript{35}Joskow’s formal modelling of profitability and leverage also preceded and thus did not reflect developments in the theory of required compensation for risk bearing by insurance company owners.

\textsuperscript{36}Zweifel and Gherli (1990) reported evidence of lower expense ratios for independent agency insurers than for exclusive agency insurers in Switzerland, but they included commission rates (which generally are higher for independent agents than for exclusive agents in the U.S. property-liability insurance market) as control variables. Their data also included experience for life and health insurance.

\textsuperscript{37}See, for example, the discussion in Doherty (1981). Moreover, Braeutigam and Pauly (1986) concluded that substantive bias in cost function estimates could arise from unobservable differences in quality that could result from price regulation.

\textsuperscript{38}Cummins and VanDerhei (1979) assumed that lower operating expenses for direct writers were prima facie evidence of superior efficiency, and concluded that regulators should take a more active role in disseminating information on prices.

\textsuperscript{39}Costly consumer search has played a role in the literature on solvency regulation for insurers (see below). Costly search associated with other dimensions of quality, such as timing and magnitude of claim payments in the absence of insurer default, also has received attention (e.g., Smallwood, 1975).
administrative, and commission expenses. Basic analysis of the relationship between insurer underwriting risk and scale of operations suggests that increasing returns to scale also could be associated with capital costs. If claim costs are not perfectly correlated across insured exposures, the standard deviation of an insurer's average claim cost will decline, *ceteris paribus*, as the number of insured exposures increases (e.g., Houston, 1964; Cummins, 1974; Venezian, 1983). If holding financial capital to reduce insurer default risk is costly (see below), this reduction in risk implies decreasing costs per insured exposure for any given probability of default because the required amount of capital per exposure will decline as the number of exposures increases. Low levels of market concentration and evidence on entry suggest that decreasing capital costs do not produce a large minimum efficient scale relative to market size. Underwriting risk declines at a decreasing rate with increases in scale, and the marginal reduction could be small relative to risk that cannot be reduced by writing more exposures (or by writing coverage in different lines of insurance).

Possible efficiency enhancing and anti-competitive aspects of institutional arrangements for pooling information among insurers have been analyzed in a number of studies (e.g., Danzon, 1983; Eisenach, 1985; also see Winter, 1988). Absent mechanisms for pooling data among insurers, claim cost forecasts might be expected to be more accurate for large firms due to their superior information. The costs of ratemaking and of complying with rate regulation also are likely to have a large fixed component. Hence, arrangements for pooling information and data analysis, some of which are made possible by the insurance industry's limited antitrust exemption under federal law and the laws of many states, are likely to reduce these costs and facilitate entry.

**Alternative Organizational Forms.** In addition to significant variation in distribution methods, insurance markets generally are characterized by a variety of organizational forms. Most important, mutual organizations commonly have a significant market share. Mayers and Smith (1981) briefly considered the ability of alternative forms of insurance company ownership to minimize the cost of conflicts between owners, policyholders, and managers (also see Fama and Jensen, 1983). Mayers and Smith argued that while mutual organization eliminates owner-policyholder conflict, it can increase the cost of controlling manager-

---

40The effects of undiversifiable risk on insurance prices are discussed below. Venezian (1984) discussed equity issues associated with insurance pricing when capital costs are additive. Much of this discussion asserted significant barriers to entry as a result of increasing returns to scale in operating costs and costly consumer search.

41This result requires that small firms are unable to infer information available to large firms by observing market prices.

policyholder conflict compared to stock organization. They predicted that mutuals will specialize in lines of insurance where managers have limited discretion to pursue their own interests at the expense of policyholders.

Mayers and Smith (1988) provided further discussion of the ability of stock, mutual, and other organizational forms used in property-liability insurance to control conflict efficiently, and they developed and tested hypotheses concerning product specialization and geographic concentration across ownership types (also see Mayers and Smith, 1986). They obtained some evidence consistent with their predictions, including significant variation in product mix across ownership types. In other analysis, Hansmann (1985) provided detailed discussion of the possible role of mutual ownership in reducing conflicts between owners and policyholders over the level of insurer default risk (also see Garven, 1987). He also considered the possible ability of mutual ownership to facilitate risk selection during the formative years of U.S. insurance markets.

**Insurance Pricing**

Economic and financial analysis of insurance pricing has largely focused on two issues: (1) the determinants of long-run equilibrium prices in view of modern financial theory, and (2) the existence and possible causes of temporal volatility in insurance prices and in the availability of coverage that cannot be explained by changes in expected costs. Both areas have important policy implications.

**Determinants of Long-run Equilibrium Prices.** Using the equilibrium risk-return relation implied by the Capital Asset Pricing Model (CAPM), Bicer and Kahane (1979) showed that equilibrium insurance underwriting profit margins (and thus premiums) were a linear function of the riskless rate of interest and the systematic risk (beta) of underwriting in the absence of income taxes. They also provided estimates of underwriting betas using accounting data for different lines of insurance (also see Cummins and Harrington, 1985). Fairley (1979) (also see Hill, 1979, and Hill and Modigliani, 1986) developed a similar model and showed that with income taxes equilibrium premiums also increased with the tax rate and the amount of financial capital invested to support the sale of insurance.

42A number of studies also have analyzed short-run determinants of prices and other aspects of insurer operations using expected profit or expected utility models of insurer decision-making (e.g., Witt, 1974; McCabe and Witt, 1980; MacMinn and Witt, 1987).
Myers and Cohn (1986) criticized the ad hoc approach used by Fairley and others to apply the CAPM to contracts with multiperiod cash flows. They proposed a discounted cash flow model that would leave insurance company owners indifferent between selling policies and operating as an investment company. Key variables affecting equilibrium premiums again included tax rates on investment income, the amount of capital invested, and the required compensation to owners for risk bearing. Kraus and Ross (1982) considered application of arbitrage pricing theory to insurance pricing using both discrete and continuous time models.

The preceding studies either ignored default risk or implicitly assumed unlimited liability for insurance company owners. Doherty and Garven (1986) analyzed long-run equilibrium premiums with limited liability using discrete time options pricing theory under conditions in which stochastic investment returns and claim costs could be valued using risk neutral valuation functions. They used numerical examples to illustrate the effects of changing various parameters. Among other implications, premiums increased and default risk declined as invested capital increased. Cummins (1988) illustrated the application of continuous time options pricing theory to calculation of risk-based premiums for insurance guaranty funds. Again, numerical examples were used to illustrate the sensitivity of premiums to changes in underlying parameters.

An important implication of research on long-run equilibrium prices is that variability in claim costs that cannot be eliminated by insurer diversification raises prices (premium loadings) for any given level of default risk and thus reduces the gains from trade in insurance markets (also see Danzon, 1984, 1985; Doherty and Dionne, 1989). Hence, undiversifiable risk provides a possible explanation of why some risks may be uninsurable in addition to the effects of adverse selection, moral hazard, and insurer sales and administrative costs.

**Price Volatility and Underwriting Cycles.** Many lines of insurance appear to be characterized by "soft" markets, in which prices are stable or falling and coverage is readily available, followed by "hard" markets, in which prices rise rapidly and the number of insurers offering coverage for some types of risk declines substantially. Popular wisdom holds that soft and hard markets occur cyclically with a period of about six years. Several studies have provided empirical evidence that reported underwriting and total operating profit margins follow a second-order autoregressive process that is consistent with a cycle (Venezian, 1985; Cummins and Outreville, 1987; Doherty and Kang, 1988; also see Smith, 1989). Interest in this area was stimulated by the liability insurance "crisis" of the mid-1980s, which was characterized by dramatic increases in premiums for many commercial liability risks and by reductions in the availability of coverage.

The traditional view of underwriting cycles by insurance industry analysts emphasizes fluctuations in capacity to write coverage. According to this view, which assumes an inelastic supply of capital, competition drives prices down until capital is depleted, insurers ultimately constrain supply in order to prevent default, and attendant increases in prices and retained earnings then replenish capital until price-cutting ensues again. Berger (1988) presented a simple model of this scenario that assumed that insurers were unable to add new capital and that pricing decisions were based on beginning of period surplus.

Several studies have questioned the existence of true cycles in prices. Cummins and Outreville (1987) considered whether cycles in reported underwriting results could simply reflect insurer financial reporting procedures in conjunction with information, policy renewal, and regulatory lags. They also provided evidence that reported operating margins follow a cyclical process for many lines of insurance in the United States and other countries. Doherty and Kang (1988) essentially argued that cycles in insurer operating results reflected slow adjustment of premiums to changes in the present value of expected future costs. However, the causes of slow adjustment and the influence of slow adjustment versus changes in costs were not clear in their analysis.

Harrington (1988) analyzed industry financial results surrounding the liability insurance crisis of the mid-1980s and discussed possible causes of the crisis including cyclical effects. This study also provided evidence that rapid premium growth in general liability insurance was associated with upward revisions in insurer loss reserves for prior years.

**This article also is included in this volume because it contains a large amount of institutional background that is useful in understanding the literature and policy debate on insurance market volatility.**
business and rapid growth in reported losses for new business. The results suggested that much of the total growth in premiums during 1980-86 could be explained by growth in expected losses and changes in interest rates (i.e., by determinants of long-run equilibrium premiums). However, premiums grew slower than discounted reported losses during the early 1980s and faster than discounted reported losses during 1985-86, a result that is consistent with cyclical effects.

McGee (1986) suggested that heterogeneous expectations of future claim costs among insurers could lead to price-cutting that characterizes soft markets. Harrington (1988) questioned whether aggressive behavior by firms with little to lose in the event of default could lead to excessive price-cutting. Winter (1988 and 1989) developed a model in which undiversifiable risk and constraints on external capital flows (such as those that might arise from asymmetric information between insurer managers and investors or from income tax treatment of shareholder dividends) and solvency (which could be imposed by regulators or reflect policyholder preferences) could lead to periods of soft markets followed by sharp increases in prices. His model predicts a negative relation between price and capital. He reported (1989) some evidence consistent with this prediction using aggregate industry data prior to the crisis of 1985-86, at which time the relationship became positive. Volatility in the commercial liability insurance market during the 1980s has led to a number of recent working papers (several of which only contain preliminary analysis and results). Subjects covered include insurer responses to exogenous shocks to capital (Gron, 1999; Cummins and Danzon, 1990), the sensitivity of premiums to interest rates (Doherty and Garven, 1990), the possible effects of regulation (Winter, 1988a; Tennyson, 1989), and possible causes of price-cutting in soft markets (Harrington and Danzon, 1990).

Insurance Regulation

Most economic analyses of regulation of insurance markets have focused on solvency regulation and regulation of premium rates and the availability of coverage. Theoretical work has had both positive and normative aspects. Most empirical work has focused on estimating the effects of regulation.

Default Risk and Solvency Regulation. Solvency regulation in the United States has three major facets: (1) direct controls on certain activities and financial reporting, (2) monitoring of insurer behavior, and (3) a system for paying claims of insolvent insurers (see Harrington and Danzon, 1986, for details). Direct controls include minimum capital requirements and limitations on investment activities. The principal monitoring system is administered by the National Association of Insurance Commissioners. Guaranty funds exist to pay claims of insolvent property-liability insurers in all states; many states have similar arrangements for other types of insurance. The traditional rationale for solvency regulation is that consumers are unable to monitor the risk of insurer default.

Actuarial literature (see Kasteijn and Rammerswaal, 1986, for a survey) has analyzed default risk as a function of various operating and financial decisions or analyzed decisions necessary to achieve a given probability of default (which generally is presumed to be chosen by regulators or management). Portfolio models of property-liability insurance company behavior (e.g., Michaelson and Goshay, 1967; Kahan and Nye, 1975; Hammond and Shilling, 1978) have either treated default risk as exogenously determined or subject to insurer choice. Economic factors that could influence this choice have not been the focus of this literature.

More recently, economic analysis of insurer default risk has focused on factors that influence insurer capital decisions under default risk. Building on the work of Borch (e.g., Borch, 1982; also see DeFinetti, 1957), Munch and Smallwood (1982) and Finsinger and Pauly (1984) model insurer default risk assuming that insurers maximize value to

---

Venezian (1985) suggested that industry wide use of suboptimal forecasting methods could produce cycles.

Winter also analyzed the implications of his model for the availability of coverage for risks with a high degree of uncertainty about future costs. Other studies that have dealt with the effects of uncertainty on availability and contract design include Danzon (1984, 1985) and Doherty and Dionne (1989). Also see Priest (1987) and Clarke, et al. (1986).

---

See Kunreuther, Kleindorfer, and Pauly (1983) for an overview of insurance regulation that also discusses compulsory insurance requirements. Possible conflicts between regulatory goals of reducing rates and promoting solvency have been discussed in many studies (e.g., Borch, 1974).

Almost all guaranty funds are financed by post-insolvency assessments on surviving insurers. The scope of coverage is limited. For example, the maximum property-liability claim payable commonly is $300,000 or less except for workers' compensation claims, which generally are fully covered.
shareholders, that demand is inelastic with respect to default risk, and that investing financial capital to support insurance operations is costly. The principal implication is that optimal capital is positively related to the amount of loss that shareholders would suffer if claim costs were to exceed the firm’s financial assets. Munch and Smallwood (1982) considered possible loss of goodwill in the event of default; Finsinger and Pauly (1984) assumed that an entry cost would be forfeited that otherwise would allow the firm to continue operating (also see Tapiero, Zuckerman, and Kahane, 1979). If shareholders have nothing to lose, they will not commit any capital. If they are exposed to loss, and if it is assumed that firms cannot add capital after claims are realized, firms will commit some capital ex ante.\textsuperscript{51}

In an empirical analysis of the effects of solvency regulation using cross-state data, Munch and Smallwood (1990) estimated the impact of minimum capital requirements and other forms of solvency regulation on the number of insurers selling coverage and the number of insolencies. While subject to significant data limitations, their results provided some evidence that minimum capital requirements reduced insolvencies by reducing the number of small domestic insurers in the market. They also compared characteristics of solvent and insolvent firms and concluded that the results were consistent with selection of default risk to maximize firm value.

Other empirical studies generally have focused on predicting insurer defaults using financial data without closely relating the variables chosen to the theory of default risk (e.g., Pinches and Trieschmann, 1973; Harrington and Nelson, 1986; McDonald, 1988). Not much is presently known about the magnitude of the effects of regulatory monitoring and guaranty funds on default risk.

Rate Regulation. Regulation of rates, which is used primarily in property-liability insurance, can affect an insurer’s average rate level or overall percentage change in its rates during a given period. It also can affect rate differentials between groups of consumers by imposing limits on voluntary or involuntary market rates for particular groups or by restricting risk classification.\textsuperscript{52}

Voluntary market rates for most U.S. property-liability lines presently are subject to prior approval regulation in about half of the states. Most states had prior approval regulation during the 1950s and 1960s, and rate regulation was likely to have encouraged insurers to use rates developed by rating bureaus (Joskow, 1973; Harrington, 1984; also see Danzon, 1983). A trend towards deregulation began in the late 1960s and continued until the early 1980s. A number of states reregulated commercial liability insurance rates following the liability insurance crisis of 1985-86. California adopted prior approval regulation for property-liability insurance with the enactment of Proposition 103 in 1988. Several additional states either have reenacted or are considering reenactment of prior approval regulation.\textsuperscript{53}

Most studies of rate regulation have estimated the impact of voluntary market rate regulation in auto insurance on average rate levels for the overall (voluntary and involuntary) market.\textsuperscript{54} Major hypotheses have been that regulation raises rates due to capture by industry, that regulation has short-run effects due to regulatory lag, and that regulation persistently reduces rates due to consumer pressure (see Harrington, 1984). Most studies have regressed either the statewide ratio of premiums to losses (or of losses to premiums) on a rate regulation dummy variable and on a variety of control variables. Harrington (1987) used this procedure and maximum likelihood estimation to provide evidence of cross-state variation in the impact of regulation. The results of this and other studies using data from the late 1970s and early 1980s (e.g., Pauly, Kleinodner, and Krunerather, 1986; Grabowski, Viscusi, and Evans, 1989) suggested that on average prior approval regulation reduced the ratio of premiums to losses.

\textsuperscript{51}The literature on capital decisions by banks contains similar results (e.g., Herring and Vankudre, 1987). Doherty (1989) and Tapiero, Kahane, and Jacques (1986) considered insurer capital decisions when demand for coverage depends on default risk. Following Meyers and Smith (1981, 1988), Garven (1987) analyzed default risk within an agency cost framework in which shareholders, managers, sales personnel, and policyholders have different incentives regarding default risk.

\textsuperscript{52}For background information on insurance rate regulation in the United States, see Harrington (1984). Involuntary markets, which are important mainly in auto, workers’ compensation, and medical malpractice insurance, include mechanisms such as assigned risk plans and joint underwriting associations. They require joint provision of coverage by insurers at a regulated rate.

\textsuperscript{53}California and a few other states also enacted rate "rollbacks" during the last few years.

\textsuperscript{54}Several studies have estimated the impact of prior approval regulation in other lines of business without firm conclusions (e.g., Stewart, 1987; Cummins and Harrington, 1987; D’Arcy, 1988; and Rizzo, 1989). It is very difficult to control for factors that could be expected to influence premiums (or the ratio of premiums to losses) for commercial lines in the absence of rate regulation. A prior, prior approval regulation is likely to have little or no impact in some commercial lines due to the widespread use of individual risk rating procedures (Stewart, 1987).
Some evidence of variation in the impact of prior approval regulation across states was provided in Harrington (1987) and several other studies, but causes of such variation generally were not addressed. A large amount of anecdotal evidence suggests that substantial regulatory intervention in insurance pricing tends to occur in states where the unregulated cost of coverage would be relatively high, that regulation favors high risk groups, and that this eventually has occurred in response to restrictive regulation. Pauly, Kleindorfer, and Kunreuther (1986) provided evidence that direct writer market share was significantly lower in states with prior approval regulation. Building on the work of Ippolito (1979), they also provided evidence that restrictive rate regulation was associated with lower operating expenses (and presumably lower quality; also see Braeutigam and Pauly, 1986).

Involuntary markets in auto insurance have been found to be significantly larger in states with prior approval regulation of voluntary market rates (e.g., Ippolito, 1979; Grabowski, Viscusi, and Evans, 1989). Involuntary market rate regulation and state restrictions on risk classification (e.g. unisex rating rules) also will affect involuntary market size (as was implied by Joskow, 1973). The relative effects of these influences and of voluntary market rate regulation would be difficult to sort out. Voluntary and involuntary market regulation of auto liability insurance rates could reduce the number of uninsured drivers by lowering rates to drivers who otherwise would fail to buy coverage (see Kunreuther, Kleindorfer, and Pauly, 1983; Keeton and Kweer, 1984). If so, the efficiency loss that otherwise would be expected from rate regulation would be mitigated. Not much is known about the magnitude of these effects or the effects of insurance rate regulation on decisions to drive and the frequency and severity of accidents.

Bibliography


