ABSTRACT

Kenneth Arrow and Karl Borch published several important articles in the early 1960s that can be viewed as the beginning of modern economic analysis of insurance activity. This chapter reviews the main theoretical and empirical contributions in insurance economics since that time. The review begins with the role of utility, risk, and risk aversion in the insurance literature and then summarizes work on the demand for insurance, insurance and resource allocation, moral hazard, and adverse selection. It then turns to financial pricing models of insurance and to analyses of price volatility and underwriting cycles; insurance price regulation; insurance company capital adequacy and capital regulation; the development of insurance securitization and insurance-linked securities; and the efficiency, distribution, organizational form, and governance of insurance organizations.

Keywords: Insurance, insurance market, risk sharing, moral hazard, adverse selection, demand for insurance, financial pricing of insurance, price volatility, insurance regulation, capital regulation, securitization, insurance-linked security, organization form, governance of insurance firms.

JEL codes: D80, D81, D82, G22, G30.

* Comments on an earlier version by L. Eeckhoudt, C. Gollier, M. Machina, and K. Viscusi were very useful.
INTRODUCTION

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; Allais, 1953a; 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.¹

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.² 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.³ Arrow (1963) showed, in

¹ Borch (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.
² Arrow (1963) is reprinted in Diamond and Rothschild (1978) and Borch (1960, 1961) are reprinted in Borch (1990).
³ In the insurance economics literature, coinsurance refers to a contract in which the insurer pays a fixed proportion of any claim amount.
absence of asymmetric information, that full insurance above a deductible is optimal when the premium contains a fixed-percentage loading. Raviv (1979) proved that convex insurance costs and risk aversion on the part of the insurer are explanations for coinsurance above a deductible in absence of asymmetric information. These last two results were extended by Blazenko (1985), Gollier (1987a) and others. Gollier (2013) offers an extensive review of this literature.

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 (Wilson, 1968; Ross, 1973; Holmstrom, 1979; Shavell, 1979a), 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, 2013).

The remainder of this chapter reviews the main developments of insurance economics subsequent to the pathbreaking work of Arrow and Borch. The remaining sections include contributions related to insurance economics that cover the following subjects: (1) utility, risk, and risk aversion in the insurance literature, (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) financial pricing models of insurance, (7) price volatility and underwriting cycles, (8) price regulation, (9) capital adequacy and capital regulation, (10) securitization and insurance-linked securities, and (11) efficiency, distribution, organizational form, and governance.

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. We do not attempt to cover the wide variety of applications of insurance

---

4 See Lemaire (1990) for a survey of these applications.
5 See Boyle (1990) for a survey of Borch's scholarly contributions.
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, and, to a lesser extent, life insurance. However, our discussion helps to illustrate issues, concepts, and methods that are applicable in many areas of insurance.\footnote{Compare, for example, the surveys of health insurance by McGuire (2012) and Morrisey (2013).}

**Utility, Risk, and Risk Aversion in the Insurance Literature**

**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 (1947), it remains the workforce in the study of optimal insurance decisions. The linear expected utility model remains the standard paradigm used to formally analyze economic behavior under uncertainty and to derive applications in many fields such as insurance (Drèze, 1974; Schoemaker, 1982; see also the recent survey of Karni, 2013). 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 $E_A U > E_B U$ (where $E_i U$ 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.\footnote{See Machina (1987) for an analysis of the limitations of the linear expected utility model. See Drèze (1987) for an analysis of the foundations of the linear expected utility model in presence of moral hazard. For analyses of the foundations and economic implications of linear state-dependent preferences, see Karni (1985), Drèze (1987), Karni (1990), and Viscusi and Evans (1990).}

Linearity in probabilities is directly associated with the independence axiom (Machina, 1987; as well the survey by Quiggin, 2013). This axiom has been challenged by many researchers, starting with Allais (1953b) 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. For contributions using non-linear models, see Karni, 1990a and Machina, 2013. The classical expected utility model remains however the most useful model for insurance analysis.
**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. They are very important to compare optimal insurance contracting by different risk averse individuals and to analyze insurance demand. They also explain the intensity of self-insurance, self-protection and precaution. Given a von Neumann-Morgenstern utility function of wealth, \( U(W) \) 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 \( \Pi_U \). The risk premium can be interpreted as the largest sum of money a risk averse decision maker with a given utility function is willing to pay above the expected loss (actuarially fair premium) to avoid a given risk. Insurers must evaluate this risk premium when they set the total insurance premium. Moreover, an insured with utility function \( U \) is said to be more risk averse than another insured with utility function \( V \) if and only if \( \Pi_U \geq \Pi_V \) when both face the same risky situation and have identical non-random initial wealth. Finally, the absolute measure of risk aversion corresponding to a given utility function \( (-U''/U') \) is said to be non-increasing in wealth, \( W \), if and only if, in the same risky situation, \( \Pi_U(W_1) \geq \Pi_U(W_2) \) for \( W_1 \leq W_2 \). A necessary condition for decreasing absolute risk aversion is that \( U'''(W) > 0 \). As we will see later, these concepts are very useful to study the properties of insurance demand and other forms of protection.

**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 \( E_XU < E_YU \) (where \( E_iU \) is the linear expected utility associated with the random variable \( i \)).

---

8 A concept of partial risk aversion also has been defined by Menezes and Hanson (1970) and Zeckhauser and Keeler (1970). See Dionne (1984) for an application to insurance economics and Briys and Eeckhoudt (1985) for other applications and a discussion of the relationships between the three measures of risk aversion.

9 See Ross (1981), Kihlstrom, Romer, and Williams (1981), and Doherty, Loubergé, and Schlesinger (1987) for analyses of risk aversion with random independent initial wealth and Dionne and Li (2012a) for the introduction of random dependent initial wealth.

10 An equivalent condition is that \( -U''/U' > 0 \) where \( -U''/U' \) is absolute “prudence” (Kimball, 1990). Prudence measures how an individual's preferences affect optimal values of decision variables such as savings. A more prudent agent should save more to protect future consumption following an increase in future income risk.
Insurance contracts with actuarially fair premiums can be interpreted in terms of a mean preserving spread since 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) corresponds to 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 on insurance demand and other optimal decisions under uncertainty (Rothschild and Stiglitz, 1971; Meyer and Ormiston, 1985; Laffont, 1989). 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. Alarie, Dionne, and Eeckhoudt (1992) 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” (Eeckhoudt and Hansen, 1980), “tail dominance” (Eeckhoudt and Hansen, 1984), a relatively strong increase in risk (Black and Bulkley, 1989) and a relative weak increase in risk (Dionne, Eeckhoudt, and Gollier, 1993). 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 by a risk averse firm under uncertainty (Sandmo, 1971; Leland, 1972), optimal saving under uncertainty (Sandmo, 1970), optimal portfolio choice (Meyer and Ormiston, 1989; Laffont, 1989), and optimal insurance decisions (Alarie, Dionne, and Eeckhoudt, 1992). See Eeckhoudt and Gollier (2013) for a general analysis of restrictions on the utility function to obtain intuitive results on insurance demand following changes in insurable risk and background risk.
DEMAND FOR INSURANCE

Basic Models of Coinsurance and Deductible Choice. Mossin, (1968) and Smith (1968) propose a simple model of insurance demand in which a risk averse decision maker has a random wealth (W) equal to W₀ – L where W₀ is nonstochastic initial wealth and L is a random insurable loss with a given distribution function. To illustrate this model, first assume that the individual can buy coverage \( \alpha L \) \((0 \leq \alpha \leq 1)\) for a premium \( \alpha P \) where \( P \equiv \lambda E(L) \), \( \alpha \) is the rate of insurance coverage (the coinsurance rate), \( \lambda (\lambda \geq 1) \) is the premium loading factor, and E(L) is the expected loss. It can be shown that the optimal insurance coverage is such that \( 0 \leq \alpha^* \leq 1 \) for a premium \( P \) where \( P \geq P \geq E(L) \) and \( P = \bar{\lambda} E(L) \) solves:

\[
E \left[ U \left( W_0 - L + \alpha^* \left( L - \bar{\lambda} E(L) \right) \right) \right] = E \left( W_0 - L \right)
\]

and where U is a von Neumann-Morgenstern utility function \((U'(\cdot) > 0, U''(\cdot) < 0)\). \( EU(W_0 - L) \) is the level of expected utility corresponding to no insurance. Hence, when the premium loading factor exceeds one but is less than \( \bar{\lambda} \), partial coverage \((0 < \alpha^* < 1)\) is demanded. The optimal coverage is obtained by maximizing \( E \left[ U \left( W_0 - L + \alpha \left( L - \lambda E(L) \right) \right) \right] \) over \( \alpha \) and the constraint that \( 0 \leq \alpha \leq 1 \). One can verify that the solution of this problem corresponds to a global maximum under risk aversion.

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

\[
U \left( W - E(L) - \Pi^U \right) = EU \left( W_0 - L \right). 
\]

A more risk averse individual with utility function V such that \( V = k(U), k' > 0, \) and \( k'' < 0 \) (Pratt, 1964) will accept to pay a risk premium \( \Pi^V \) greater than \( \Pi^U \). Finally, when \( \lambda > \bar{\lambda} \), a risk averse individual does not buy any insurance.

---

11 In this section, we limit discussion to the case where insurance premiums are exogenously determined. The general case is considered in the next section.
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 \( \lambda \) or the price of insurance increases: a negative substitution effect and a positive wealth effect. Hoy and Robson (1981) propose an explicit theoretical condition under which insurance is a Giffen good for the class of constant relative risk aversion functions. Briys, Dionne, and Eeckhoudt (1989) generalize their analysis and provide a necessary and sufficient condition for insurance not to be a Giffen good. This condition bounds the variation of absolute risk aversion so that the wealth effect is always dominated by the substitution effect.

The demand for insurance can also be affected by the level of risk. Alarie, Dionne and Eeckhoudt (1992) present sufficient conditions to obtain the intuitive result that a risk averse insured will increase his demand for insurance (coinsurance \( \alpha \)) when a mean preserving increase in risk is introduced in the initial loss distribution. Finally, the level of \( \alpha^* \) can be affected by the level of risk aversion. As discussed by Schlesinger (2013), an increase in risk aversion will lead to an increase in coinsurance, at all levels of wealth.

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

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

where \( D^* \geq 0 \) is the optimal deductible and \( \bar{P} = \lambda \bar{E}(I(L)) \). 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 is proved by Schlesinger (1981) for the standard EU(W) model. Moreover, under decreasing absolute risk aversion, \( dD^*/dW_0 > 0 \) (Mossin, 1968). It is possible to infer the degree of risk aversion of insurance buyers by observing their choices of deductibles (Drèze, 1981; Cohen and Einav, 2007). All these results are obtained under the assumption the insurer is not exposed to any solvency risk. With solvency...
risk the above results do not in general hold (Doherty and Schlesinger, 1990). For example, full insurance coverage is no more optimal when \( \lambda = 1 \). Moreover, an increase in risk aversion does not necessarily lead to a higher level of insurance coverage. Consequently, insurance is not necessarily an inferior good under solvency risk.

**Optimal Coverage with Random Initial Wealth.** If \( W_0 \) is an uninsurable random variable rather than fixed, the optimal level of coverage \((\alpha^{**})\) depends on the statistical relationship between \( W_0 \) and \( L \). Let us suppose that \( W_0 \) and \( L \) are independent random variables. This case is often described as the independent background risk case. One can show that, if \( U(W) \) is quadratic or corresponds to constant absolute risk aversion, the presence of a random \( W_0 \) does not affect the optimal level of insurance chosen.

If we assume that the correlation coefficient is a sufficient measure of the dependence between \( W_0 \) and \( L \), Doherty and Schlesinger (1983b) show that the Mossin (1968) and Smith (1968) result on the optimal coinsurance rate \((\alpha^*)\) with fixed \( W_0 \) is qualitatively similar to the case in which \( W_0 \) and \( L \) are independent random variables \( \alpha^{**} \). That is, \( \alpha^{**} = 1 \) when the premium is actuarially fair and \( \alpha^{**} < 1 \) when \( \lambda > 1 \). Moreover, Eeckhoudt and Kimball (1992) show that \( \alpha^{**} \neq \alpha^* \) when \( \lambda > 1 \). Specifically, they show that \( \alpha^{**} > \alpha^* \) under standard risk aversion (decreasing risk aversion and decreasing prudence). This result was proved for independent risks. They also analyze optimal deductibles and show, under the same conditions, that \( 0 < D^{**} < D^* \) where \( D^{**} \) is the optimal deductible when \( W_0 \) and \( L \) are independent random variables and \( D^* \) is the optimal deductible with fixed \( W_0 \). Hence, with independent risks, more coverage is demanded than with fixed wealth in both coinsurance and deductible contracts, under standard risk aversion and non-actuarial insurance pricing.

It was mentioned above that a more risk averse individual with utility \( V \) is willing to pay a higher risk premium for full insurance than a less risk averse individual with utility \( U \) when \( W_0 \) is not random. This result also holds when \( W_0 \) and \( L \) are independent random variables. For example, Kihlstrom, Romer, and Williams (1981) show that a more risk averse individual with utility \( V \) is willing to pay a higher premium than an individual with utility \( U \) if the absolute risk aversion for either individual for realized levels of \( W_0 \) is non-increasing in wealth.

If \( W_0 \) and \( L \) are negatively (positively) correlated, high losses are likely to accompany low (high) values of \( W_0 \). Doherty and Schlesinger (1983b) show in the case of a two-state 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), Schu-elenburg (1986), Turnbull (1983), Eeckhoudt and Kimball (1992) and Lévy-Garboua and Montmarquette (1990).} Eeckhoudt and Kimball (1992) show, in presence of a positive relationship between $W$ and $L$, that decreasing absolute risk aversion and decreasing absolute prudence are sufficient to guarantee that the presence of a background risk increases the optimal insurance coverage against $L$. To obtain this result, a simple positive correlation between $L$ and $W$ is not sufficient. The authors use the stronger assumption that “the distribution of background risk conditional upon a given level of insurable loss deteriorates in the sense of third-order stochastic dominance as the amount of insurable loss increases” (p. 246). In conclusion, there is no general measure of dependency that can yield general results on insurance demand in presence of dependent risk. Aboudi and Thon (1995) and Hong et al (2011) propose some orderings such as expectation dependence to obtain intuitive results (see also Li, 2011, and Dionne and Li, 2012b, for more general discussions on expectation dependence and Schlesinger, 2013, for a longer discussion on the effect of dependent risk on insurance demand).

**Insurance, Portfolio Choice, and Saving.** Mayers and Smith (1983) and Doherty (1984) analyze the individual demand for insurance as a special case of a general portfolio choice model. They introduce nonmarketable assets (such as human capital) and other indivisible assets (such as houses) in a capital asset pricing model to simultaneously determine the demand for insurance contracts and the demand for other assets in the individual portfolio.\footnote{See Kahane and Kroll (1985) and Smith and Buser (1987) for extensions of these models.} Mayers and Smith (1983) propose a sufficient condition for a separation theorem between insurance decisions and other portfolio decisions. Their condition is that the losses of a particular type are orthogonal to the insured’s gross human capital, the payoff of all marketable assets, and the losses of other insurable events. This is a strong condition and their analysis suggests that portfolio and insurance decisions are generally 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) for dependent risks.

Moffet (1975, 1977) and Dionne and Eeckhoudt (1984) provide joint analyses of the saving (consumption) and insurance decisions in a two-period model. Dionne and Eeckhoudt (1984)
generalize Moffet’s results and show that, under decreasing temporal risk aversion, savings 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 present two conditions under which a separation theorem holds between insurance and saving decisions: 14 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 consider a menu of risky assets while Dionne and Eeckhoudt (1984) consider only one risky 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).

Briys (1988) extends 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 loss (i.e., L'(y) < 0 with L"(y) > 0) while self-protection refers to actions (x) that reduce the probability p(x) (frequency) of accidents (p'(x) < 0 with p"(x) > 0) in a two-state environment. 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. 15

While Ehrlich and Becker (1972) focus on the interaction between market insurance and activities involving either self-insurance or self-protection, they do not study in detail interactions between self-insurance and self-protection with and without the existence of market insurance.

---

14 See Drèze and Modigliani (1972) for another sufficient condition on utility to obtain separation between consumption, portfolio, and insurance decisions.

15 See Winter (2013) for an analysis of self-protection and self-insurance under asymmetric information. We will come back on this issue in the moral hazard section of this chapter.
Boyer and Dionne (1983, 1989) and Chang and Ehrlich (1985) present 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) show, in a model with two states of the world, that this is not always the case: more risk averse individuals may undertake less self-protection.\(^{16}\) Briys and Schlesinger (1990) document that self-insurance corresponds to a mean preserving contraction in the sense of Rothschild and Stiglitz (1970) while self-protection does not necessarily reduce the risk situation. Chiu (2005) did a further step by showing that a mean-preserving increase in self-protection is a special case of a combination of an increase in downside risk and a mean-preserving contraction. This introduces the role of prudence in the comparative static analysis of self-protection. In fact, Dionne and Li (2011) obtain that a risk averse individual will produce more self-protection that a risk neutral individual if \(p(x) < 1/2\) and if prudence is lower than an upper bound. This upper bound can be interpreted as the skewness of the loss distribution per amount of loss for risk averse agents with decreasing prudence.

**Precaution.** Precaution differs from protection in many ways. Courbage, Rey and Treich (2013) propose this analytical difference: protection is a static concept while precaution is dynamic, in the sense that precaution evolve with the observation of more information. In fact, in practice, precaution activities are implemented because decision makers do not have enough information to implement self-protection, self-insurance and insurance decisions. Dachraoui et al (2004) and Chiu (2005) show how self-insurance and self-protection are related to willingness to pay (Drèze, 1962; Jones-Lee, 1974) for reducing the amount of loss or the loss probability, respectively. When precaution decisions are made, this is generally because decision makers suspect a potential risk but cannot attribute a probability to this risk nor an evaluation of the amount of loss. Precaution has been associated to the irreversibility effect (Henry, 1974; Arrow and Fisher, 1974; Epstein, 1980).

Although the analysis of Epstein (1980) did provide a strong reference framework to analyze how new information over time would improve precaution decision making, there are not clear

\(^{16}\)See Hiebert (1989), Briys and Schlesinger (1990), Julien et al (1999), Dachraoui et al (2004), and Dionne and Li (2011) for extensions of their analysis. A recent literature review on this subject is presented in Courbage, Rey and Treich (2013).
results in the literature on how risk averse individuals make decisions about precaution. One explanation offered by Courbage, Rey and Treich (2013) is that the Blackwell notion of better information used in this literature is too general to obtain clear conclusions. More research is necessary on this important concept because our societies face many situations where the scientific knowledge is not developed enough to make informed decision even if some decisions must be made in the short run. Climate changes and associated hurricanes are evident examples; pandemics are other examples. The precautionary principle is a concept that may help to improve research in that direction (Gollier et al, 2003).

**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 them in whole or in part. However, there are other commodities for which market substitutes are imperfect or do not exist. Examples include good health, the life of a child, and family heirlooms. For these “commodities”, an accident generates more losses than monetary losses: it 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 \( U_i^I(W) > U_i^II(W) \) for all \( W \) (where \( U_i^i(W), i = I, II, \) is the utility in state \( i \)). With a monetary loss (\( L > 0 \)), \( U_II(W) - U_II(W_0 - L) \) measures the disutility of the monetary loss in the accident state and \( U_I(W_0 - L) - U_II(W_0 - L) \) measures the disutility of the non-monetary loss from state I to state II.

Marginal utility of wealth also depends on the state of the world. Three cases are usually considered: (1) \( U_w^I = U_w^{II} \) for all \( W \); (2) \( U_w^I > U_w^{II} \) for all \( W \); and (3) \( U_w^I < U_w^{II} \) for all \( W \) where \( U_w^i \) denotes \( \partial U_i^i / \partial W \). It can be shown that \( \alpha * \frac{U_w^I}{U_w^{II}} < 1 \) for a policy with an actuarially fair premium as long as \( U_w^{II} > U_w^I \) for all \( W \). That is, 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 \( W \). Karni (1985) show 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 (Dionne and Ingabire, 2001).

**Corporate Demand for Insurance and Enterprise Risk Management.** Portfolio decisions have implications for the demand of 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 corporate demand for insurance. Specifically, if shareholders can costlessly eliminate the risk of corporate losses in their own portfolio’s 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.¹⁷ Mayers and Smith (1982) analyze the corporate demand for insurance in the perspective of modern finance theory (also see Main, 1982; Mayers and Smith, 1990; MacMinn and Garven, 2013). Market imperfection can explain corporate demand for insurance. They discuss how solvency costs; risk aversion by stakeholders such as managers, employees, customers, and suppliers; efficiencies in claims administration by insurers; and a number of other factors such as taxes or investment financing in imperfect capital markets, each can provide an incentive for the purchase of insurance (or risk hedging) even when shareholders can costlessly eliminate risk through portfolio diversification. In a later study, Mayers and Smith (1987) consider the possible ability of insurance to increase shareholder wealth by mitigating the underinvestment problem that was originally analyzed by Myers (1977). This literature is now related to that of firms’ risk management (Stulz, 1990; Tufano, 1996; Cummins et al, 2000; Dionne and Triki, 2011; Hoyt and Liebenberg, 2011; Campello et al, 2011).

Although perfect markets finance theory provides little rationale for widely held firms to expend resources to hedge unsystematic risk, various market imperfections create opportunities for such firms to maximize market value through hedging. The principal market imperfections that motivate corporate hedging are corporate income taxation (Smith and Stulz, 1985; Graham and Smith, 1999; Graham and Rogers, 2002), financial distress costs (Smith and Stulz, 1985), investment opportunity costs (Froot, Scharfstein, and Stein, 1993; Froot and Stein 1998), information asymmetries (DeMarzo and Duffie, 1991), and corporate governance considerations

¹⁷ This statement also holds if insurable risk has an undiversifiable (i.e., market) component, since insurers have no comparative advantage in bearing market risk (see Main, 1982).
(Dionne and Triki, 2011). Firms also engage in hedging for non-value-maximizing reasons such as managerial risk aversion (Stulz, 1990; Tufano, 1996).

In recent years the literature on corporate risk management has begun to emphasize unified management of all major risks confronting organizations through a process of “enterprise risk management.” In principle, enterprise risk management has the potential to increase value by considering correlations among distinct risks and focused more attention on firms’ aggregate exposure to loss, thus reducing, for example, redundancies in hedging and insurance coverage (e.g., Harrington and Niehaus, 2002). Another question relates to the value of risk management to the shareholders’ wealth.18

**Insurance and Resource Allocation**

Allais (1953a) and Arrow (1953) proposed general equilibrium models of resource allocation in the presence of uncertainty at a meeting on the subject in Paris during 1952. Debreu (1953) extended Arrow’s (1953) contribution to a general framework of resource allocation under uncertainty.19 In this framework, physical goods are redefined as functions of states of the world and a consumption plan specifies the quantity of each good consumed in each state. Preferences among consumption plans reflect tastes, subjective beliefs about the likelihoods of states of the world, and attitudes towards risk.20 However, beliefs and attitudes towards risk do not affect producer behavior since for given contingent prices; there is no uncertainty about the present value of production plans. The existence of a competitive equilibrium that entails a Pareto optimal allocation of goods and services can be demonstrated for this economy.

Insurance markets can be viewed as markets for contingent goods. Borch (1962) proposed the first formal model of optimal insurance contracts. He presented a very elegant comparison between a general model of reinsurance and the Arrow-Debreu model with pure contingent

---

18 Hoyt and Liebenberg (2011) obtain a positive relation between firm value and enterprise risk management for US insurers. Cummins, et al (2009) argue that risk management and financial intermediation are two activities that may be used by insurers to improve efficiency, where efficiency is gauged by the capacity to reduce the costs of providing insurance. They measure insurer efficiency by estimating an econometric cost frontier. Because risk management and financial intermediation are key activities for insurers, they treat these activities as endogenous. Their econometric results suggest that risk management significantly increases the efficiency of U.S. property/casualty insurance industry.

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

20 In the Arrow-Debreu world each agent has incomplete information about states of nature (uncertainty) but all agents share the same information (Radner, 1968). The latter assumption rules out moral hazard and adverse selection problems.
goods and contingent prices for every state of the world. As noted earlier, Borch’s insurance model can be reinterpreted in terms of standard insurance contracts. Two of his major contributions were to provide conditions for Pareto optimal exchange of risk and to show how risk aversion by insurers can explain partial coverage. Arrow (1965) used the same argument to introduce some element of coinsurance in optimal insurance contracts. Moreover, Arrow (1963) showed that if a risk neutral insurer offers a policy with a premium equal to the expected indemnity plus a proportional loading then the optimal contract provides full coverage of losses above a deductible. These forms of partial insurance limit the possibilities of risk shifting between economic agents (Arrow, 1965).

Raviv (1979) extended these results and showed that a Pareto optimal contract involves both a deductible and coinsurance of losses above the deductible. He also showed that the optimal contract does not have a deductible if the administrative cost of 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 insurance 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 IN INSURANCE MARKETS

The concept of moral hazard was introduced in the economics literature by Arrow (1963) and Pauly (1968) (see also Kihlstrom and Pauly, 1971; and Spence and Zechauser, 1971). Two types of moral hazard have been defined according to the timing of an individual’s actions in relation to the realization of the state of nature. They are identified as ex ante and ex post moral

---

22 See Gollier (1987a) for an extensive analysis of this constraint and Gollier (1992) for a review of optimal insurance contracting.
23 See Rowell and Connelly (2012) for a historical analysis of the origin of moral hazard.
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 (1976a), Holmstrom (1979), and Shavell (1979a, 1979b) consider 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 of risk averse individuals to take care compared to the case of full information. Thus, there is a trade-off between risk sharing and incentives for prevention when the insured (or the agent) is risk averse.

Shavell (1979b) 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 or prevention) to show that partial insurance coverage is optimal in the presence of moral hazard. He emphasized that the type of care has a major impact on the optimal solution. As the cost of care decreases from high levels to lower levels, partial coverage becomes desirable. In other words, when the cost of care is very high, partial coverage has no effect on prevention and reduces protection. Another important result is that moral hazard alone cannot eliminate the gains of trade in insurance markets when an appropriate pricing rule is implemented by the insurer (i.e., moral hazard 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 by the insurer (also see Holmstrom, 1979).

Shavell’s two-state model did not permit a detailed characterization (security design) 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, 2013, for detailed analysis). For example, when prevention affects only the accident probability (and not the conditional loss distribution), a fixed deductible is optimal.

24 The ex ante actions can affect event probabilities, event severity, or both (see Winter, 2013, for more details).

25 Also see Pauly (1974) for a similar model. See Dionne (1982) for a model with state-dependent preferences. It is shown that moral hazard is still an important problem when preferences are not limited to monetary losses. Viau roux (2011) extends the principal-agent model to introduce taxes.
However, when prevention affects the conditional distribution of losses, the coassurance coverage becomes necessary in order to introduce more prevention against high losses.

Moral hazard in insurance can be analyzed within a general principal-agent framework (Ross, 1973; Holmstrom, 1979; Grossman and Hart, 1983; Salanié, 1997; Laffont and Martimort, 2002; Bolton and Dewatripont, 2005). However, certain conditions must be imposed to generate predictions and obtain optimality when using the first-order approach. 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 (1985a) 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 (CDFC) and (2) the likelihood ratio has to be monotone (MLRP condition). 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.26 He then supplied an alternative set of conditions including restrictions on the agent’s utility function to validate the first-order approach (see Winter, 2013, and Arnott, 1992, for further discussion on insurance applications).

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 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.27

Long-term contracts between principals and agents can increase welfare in the presence of moral hazard (Rogerson, 1985b; Radner, 1981; Rubinstein and Yaari, 1983; Henriet and Rocket, 1986). In multiperiod insurance models, an individual’s past experience eventually gives a good

26 See LiCalzi and Speater (2003) for an extension of this research.
27 See Dye (1986) and Mookherjee and Png (1989) for recent applications of Grossman and Hart’s model.
approximation of care. Hence insurers use the individual’s past experience to determine premiums and to increase incentives for exercising care. For example, Boyer and Dionne (1989a) show empirically how past demerit points approximate drivers’ effort. In automobile insurance, insurers use the bonus-malus scheme to obtain more incentives for safe driving (see Dionne et al, 2013, for a recent survey of empirical studies on asymmetric information models in the literature on road safety and automobile insurance).

Moral hazard may alter the nature of competitive equilibrium by, for example, introducing non-convexities 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 theorem of welfare economics holds 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 (Arnott and Stiglitz, 1990; Arnott 1992).

Moral hazard also can affect standard analyses of government responses to externalities. An important example involves liability rules and compulsory insurance. With strict liability and risk averse victims and injurers, 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 assumes 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 insurance could reduce care, and it is even possible that prohibiting insurance coverage could improve the level of care.

---

28 See Harrington and Danzon (2000a) for a survey on the demand and supply of liability insurance and Shavell (2004) for a comprehensive review of the economic analysis of law. For a recent comparison of strict liability and a negligence rule for risk-incentives tradeoff, see Fagart and Fluet (2009).
The level of activity or risk exposure must also be considered in the analysis. Liability rules affect not only the level of care but also the level of activity. For example, negligence rule may continue to implement the social level of prevention, but it induces injurers to inflate their level of activity (miles driven) (see Shavell, 2004, for a complete coverage of different liabilities rules in presence of insurance).

Litigation is important in insurance economics. Litigation procedures and their costs affect the legal system’s capacity to obtain appropriate incentives and compensations. A question analysed by Shavell (2004) is the following: Why should insurers be motivated to influence litigation? First they receive the insurance compensation paid to the victims. Moreover, defendants may own liability insurance policies and insurers may have incentives to defeat plaintiffs. There may be conflicts of interest between the insurers and their clients. For example, the plaintiff’s insurer may be more willing to settle than the plaintiff and the defendant’s insurer may be less willing to settle than the defendant. As observed by Shavell (2004), “insurance generally reduces incentives to spend at trial” (p. 440) because insurers bear legal costs and often make litigation decisions. But this is not necessarily inefficient. Some forms of insurance contracts may be jointly beneficial in terms of litigation.

Finally, point-record driver’s licenses may complement insurance incentive schemes for road safety, particularly when fines for traffic violations are bounded. Bounded fine exist for different reasons: 1) many offenders are judgment proof and are unable to pay optimal fines; 2) many drivers are expected to escape from paying tickets issued by the authorities when fines are very high; 3) society thinks it is unfair that rich and reckless drivers will pay high fines and continue to drive dangerously (Shavell, 1987a, 1987b). However, fines do reinforce the effectiveness of the point record mechanism by providing normal drivers with more incentives (Bourgeon and Picard, 2007). For a recent comparison of strict liability and a negligence rule for risk-incentives tradeoff, see Fagart and Fluet (2009). See Dionne, Pinquet et al (2011) for a comparative analysis of different incentive schemes for safe driving under moral hazard. See also Abbring et al (2003, 2008) for empirical analyses of the presence of moral hazard in automobile insurance markets, and Chiappori and Salanié (2013) for a survey of empirical research on asymmetric information problems in insurance markets.

**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 action 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. Marshall (1976b), 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 problem is auditing. He shows that a deductible contract can be optimal under certain circumstances since it reduces the cost of audit.

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. Lacker 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.29

Insurance fraud (Derrig, 2002) is directly related to ex post moral hazard. The economic analysis of insurance fraud is not limited to illegal activities. The notion of fraud covers “unnecessary, unwanted, and opportunistic manipulations” (Derrig, 2002) of the insurance system. It results from asymmetric information in the claiming behavior. The recent literature on the economic analysis of insurance fraud in divided in two main approaches: that related to the costly state verification analysis of moral hazard and that related to the costly state falsification. Under costly state verification, the insurer can verify the nature of the claim at a cost. Claims verification can be deterministic or random and can be conditioned on signals received by insurers. Under costly state falsification, the insured uses resources for modifying his claim (as for a build-up). Pierre Picard (2013) proposes a general theoretical model that relates the main contributions in this literature (Picard, 1996; Crocker and Morgan, 1997; Bond and Crocker, 1997; Picard, 2000; Fagart and Picard, 1999).

A general conclusion from Picard’s survey (2013) is that insurance fraud affects the design of insurance contracts. In some cases, a deductible can be optimal but usually some degrees of coinsurance are necessary, particularly when policyholders can manipulate audit costs or falsify claims. Another issue is related to the insurers’ incentives of reducing insurance fraud. There is a free riding issue in competitive markets: Individual insurers may simply increase their premiums following an increase in ex post average losses. Some authors have suggested the development of common agencies to develop common data sets and models of fraud detection (Derrig, 2002). Finally, services providers (garage owner, doctor, lawyer…) payoffs are also affected by the insurance coverage of the service cost they provide. Preventing collusion between insured and service provider is still an open question in the literature. See Dionne and Gagné (2001, 2002), Dionne et al (2009), and Dionne and Wang (2012) for recent empirical analyses of insurance fraud in different insurance markets.

**Adverse Selection in Insurance Markets**

Adverse selection occurs in insurance markets 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 analyzed the ability of partial insurance coverage, experience rating, and risk categorization to reduce the negative effects of adverse selection.\(^{30}\) Others have considered the effect of adverse selection on economic welfare and market equilibrium.

**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 Rothschild and Stiglitz (1976), insurers offer a menu of policies with different prices and quantities so that different risks choose different insurance contracts. These strategies have been studied for single

---

\(^{30}\) We 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 signaling models.
vs. multi-period contracts, for competition vs. monopoly, and, when assuming competition, for several different equilibrium concepts.31

In a single period model with competition, Rothschild and Stiglitz (1976) show 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 reveal hidden information.32 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. Another characteristic of their definition of Nash equilibrium under adverse selection is that no contract outside the equilibrium can make a positive profit with a risk type. However, when there exist relatively few high risk persons in the market, they showed that neither a separating nor a pooling equilibrium exists.

Other equilibrium concepts that eliminate the non-existence problem have been proposed. Wilson (1977), Miyasaki (1977), and Spence (1978) considered the case in which firms anticipate that other insurers’ policies that become unprofitable as a result of new offerings will be withdrawn.33 A Wilson-Miyasaki-Spence (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.34 Finally, a WMS equilibrium is always second best efficient.

---

31 See Cooper and Hayes (1987), Crocker and Snow (1985), and Cresta (1984) for an introduction to these models and Dionne, Fombaron and Doherty (2013) for a survey on adverse selection in insurance contracts.

32 A 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.

33 The 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).

34 Each of these models either explicitly or implicitly assumes that insurers could enforce the requirement
Dahlby (1983) provided some empirical evidence of adverse selection in the Canadian automobile insurance market. He suggests that his empirical results are consistent with the WMS model with cross-subsidization between individuals in each class of risk. However, Riley (1983) argues that Dahlby’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. More recently, Chiappori and Salanié (2000) and Dionne, Gouriéroux, and Vanasse (2001) proposed a test for the presence of residual asymmetric information in insurance markets. Their test cannot separate moral hazard from adverse selection because both information problems predict a positive correlation between individual risk type and realized accidents. The separation of the two problems can be obtained by applying a causality test where contract choice predicts risks (moral hazard) and risks predict contracts choices (adverse selection). Dionne, Michaud and Dahchour (2011) propose such a test.35

**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.36 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 et al, 2013, for a more detailed review).

Dionne (1983), Dionne and Lasserre (1985), and Cooper and Hayes (1987) extend Stiglitz’s monopoly model (1977) to multi-period contracts. Dionne (1983) considers infinite length contracts without discounting while Cooper and Hayes (1987) mainly dealt with a finite horizon model (without discounting). While findings in both cases suggest that experience rating induces sorting or risk disclosure, the analyses differ in many respects. In Dionne (1983), a simple

---


36 See Dalhby (1990), Dionne and Lasserre (1987), and Dionne and Vanasse (1992) for analyses of experience rating when moral hazard and adverse selection are present simultaneously.
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 serves 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.37

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

Nilssen (2000) analyze experience rating contracts without commitment by insurers in a competitive market. His results differ from those of Cooper and Hayes and are quite similar to those of Kunreuther and Pauly (1985), who assume that insurers sell price-only policies (Pauly, 1974) rather than price-quantity policies. Another important assumption in Kunreuther and Pauly’s model is 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) provide some empirical evidence that is consistent with Kunreuther and Pauly’s model while Dionne and Doherty (1994) consider the semi-commitment model. Dionne and Doherty (1994) do not reject the presence of adverse selection.

Risk Caterogization.38 In most types of insurance, insurers classify risks using many variables. In auto insurance, for example, empirical evidence indicates that driver age and sex are significantly related to accident probabilities (Dionne and Vanasse, 1992). In particular, evidence

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

38 We limit our discussion to exogenous categorization of risks. See Bond and Crocker (1991) for an analysis of endogenous categorization of risks.
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, 2013; also see Hoy, 1982; and Rea, 1987, 1990) show 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. More recently, Rothschild (2011) showed that categorical pricing bans are inefficient even when categorization is costly. Crocker and Snow (1986) also consider the existence of a balanced-budget tax-subsidy policy that provides private incentives to use risk categorization. With appropriate taxes, they show that no agent would lose from classification. In their 1986 article, the results were obtained 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 (e.g., age and sex classification in auto insurance). Browne and Kamiya (2012) show that a policy requiring insurance buyers to take an underwriting test can provide full coverage when perfect classification is possible or when there is no residual asymmetric information in insurer’s risk classes.

**Multidimensional Adverse Selection.** The recent literature on adverse selection added other dimensions than risk to the asymmetric information between insurer and insured. For example, Villeneuve (2003) and Smart (2000) consider differences in risk aversion that are not observable by the insurer. Wambach (2000) incorporates heterogeneity in initial wealth. Snow (2009) show that profitable contracts obtained in the above contributions cannot be sustained as a Nash equilibrium when insurers offer menus of contracts. Other extensions analyze insurance contracting with multiple risks (Fluet and Pannequin, 1997; Crocker and Snow, 2011).

**Participating Contracts and Adverse Selection.** Participating contracts is a form of mutualisation where individuals’ premiums are function of individual and collective loss
experience of the insurance pool. In a recent paper, Picard (2009) shows that allowing insurers to offer participating policies in addition to non-participating ones guarantees the existence of an equilibrium in the Rothschild and Stiglitz (1976) environment. The presence of participating contracts dissuades other insurers to offer contracts to low risk individuals only as an alternative. Picard (2009) shows that when there is no Rothschild and Stiglitz (1976) pooling equilibrium, equilibrium with cross-subsidized participating contracts exists.

Other papers in the literature that suggest the coexistence of stock and mutual contracts are related to aggregate risk (Doherty and Dionne, 1993), adverse selection (Ligon and Thistle, 2005), and insolvency of stock insurers (Rees, Gravelle, and Wambach, 1999; Fagart, Fombaron, and Jeleva, 2002).

FINANCIAL MODELS OF INSURANCE PRICES

Basic financial theory predicts that in long-run equilibrium competitively determined insurance premiums, commonly known as fair premiums, will equal the risk-adjusted discounted value of expected cash outflows for claims, sales expenses, income taxes, and any other costs, including the tax and agency costs of capital. A large literature has considered the technical details in the context of specific financial pricing models.

**Basic CAPM and Related Models.** Early work on financial pricing of insurance examines the implications of the Capital Asset Pricing Model (CAPM). Biger and Kahane (1978) shows in that context and the absence of taxes that equilibrium insurance underwriting profit margins were a linear function of the riskless rate of interest and the systematic risk (beta) of underwriting. They also provide estimates of underwriting betas using accounting data for different lines of insurance (also see Cummins and Harrington, 1985, and more recently, Cummins and Phillips, 2005). Fairley (1979) develops a similar model and shows that with income taxes fair premiums increase with the tax rate and the amount of financial capital invested to support the sale of insurance. Kraus and Ross (1982) consider application of arbitrage pricing theory to insurance pricing using both discrete and continuous time models. Myers and Cohn (1986) propose 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 include tax rates on investment income, the amount of capital invested, and the required compensation to owners for risk bearing.
**Option-Based Models.** The financial pricing literature next advanced to analyzing the effects on fair premiums of limited liability and the insurer’s resulting default put. Doherty and Garven (1986; also see Cummins, 1988) analyze fair premiums with limited liability using discrete time options pricing theory under conditions in which stochastic investment returns and claim costs can be valued using risk neutral valuation, showing numerically that premiums increase and default risk declines as invested capital increases.39 A number of studies subsequently provided empirical evidence consistent with this intuitive prediction (Sommer, 1996; Phillips, Cummins, and Allen, 1998; Cummins, Phillips, and Lin, 2006).40

**Pricing with Multiple Lines of Business.** The early pricing studies essentially model prices for a company writing a single line of business. In a multiline context, a key question is whether differences in risk and capital across lines of business imply differences in prices across lines for a given level of the default put. Phillips, Cummins, and Allen (1998) extend option pricing methods to a multiline context assuming no frictional costs, such as tax and agency costs, of holding capital. They assume that in the event of default, the excess of liabilities over assets would be allocated among policyholders in proportion to their ex ante expected claims. Under this assumption, the ratio of expected default costs to expected claims is constant across lines of business, implying a constant ratio of premiums to expected claims. As a result, their model implies that it is not necessary to allocate an insurer’s capital across lines of business to determine prices.

Other studies consider by line pricing and capital allocation in the presence of frictional costs of holding capital. In this case, the question arises as to whether the frictional costs must be allocated across lines of business to determine line-specific prices. Myers and Read (2001, also see Merton and Perold, 1993) analyze a linear capital allocation rule implied by an Euler equation for the insurer’s default risk (the default put value). They derive the marginal capital increase needed from an increase in a particular source of risk to maintain a constant overall risk level. The resulting marginal changes satisfy an “adding up” property: the sum of the marginal values

---

39 Borch (1974) obtains a qualitatively similar result assuming limited liability and expected utility maximization by insurers.

40 Bauer, Phillips, and Zanjani (2013) provide an overview of more rigorous theoretical models that consider insurance prices with frictionless securities markets and no arbitrage opportunities, incomplete markets for insurance, varying degrees to which insurance risk is diversifiable, and possible dependencies between insurance and financial markets.
over different sources of risk (lines of business) equals the firm’s total capital. Subsequent work questions the risk measure chosen by Myers and Read and proposes a variety of alternative risk measures and conceptual frameworks for determining unique capital allocations (see, for example, Sherris, 2006; Grundl and Schmeiser, 2007; Zanjani, 2010; Ibragimov, Jaffee, and Walden, 2010; and the review by Bauer and Zanjani, 2013). In related work, Froot and Stein (1998) and Froot (2007) consider pricing, capital allocation, capital budgeting, and risk management decisions for financial institutions in general and insurers / reinsurers in particular.

The necessity and details of capital allocation notwithstanding, a key implication of the literature on multiline pricing is that frictional costs of holding capital will require higher prices for riskier lines of business. Particular attention has been paid to prices of catastrophe coverage. Zanjani (2002b), for example, develops a model in which prices are high for capital intensive lines of business, such as catastrophe coverage. Harrington and Niehaus (2003) explore in detail the tax costs of capital that arise from double taxation of investment income from capital and illustrate that the effects on premiums can be very large for catastrophe risk, where expected claim costs are low in relation to the amount of capital needed to achieve a low level of default risk.41 Phillips, Cummins, and Lin (2006) provide evidence that by-line insurance prices reflect capital allocations and default risk.

**Price Volatility and Underwriting Cycles**

Abstracting from technical details, financial pricing models imply that fair premiums and loss ratios (ratios of incurred losses to premiums) should vary over time in relation to the fundamental determinants of prices.42 The “fundamentals” include predicted claim costs and underwriting expenses, riskless interest rates and any systematic risk of claim costs and associated market risk premia that affects the discount rate for expected claim costs, and the tax and agency costs of holding capital to bond an insurer’s promise to pay claims. Not surprisingly, there is abundant evidence that changes in claim costs, which should be highly correlated with insurer forecasts when policies are priced, explain much of the time series variation in premiums. Examples include studies of premium growth in automobile insurance (Cummins and Tennyson, 1992) and medical malpractice insurance (Danzon, 1985, Neale, Eastman, and Drake, 2009). Numerous

---

41 Froot (2001) and Froot and O’Connell (2007) provide empirical evidence of high prices for catastrophe coverage and consider alternative explanations.

42 This and the following sections draw from Harrington, Niehaus, and Yu (2013).
studies also provide evidence, albeit sometimes weak, of the predicted inverse relationship between interest rates and loss ratios or combined ratios (e.g., Smith, 1989, Choi and Thistle, 2000, Harrington, Danzon, and Epstein, 2008; Jawad, Brueneau, and Sghaier, 2009). Other evidence and popular wisdom regarding insurance underwriting cycles are more difficult to reconcile with plausible changes in pricing fundamentals.

**Aggregate Time Series Studies.** Because data on average premiums per exposure generally are not available to researchers, most empirical analyses of price volatility in insurance markets use data on loss ratios or “combined” ratios (loss ratios plus administrative expense ratios) to control for scale effects and abstract in part from the effects of changes in claim cost forecasts over time. Many studies document empirical regularities in underwriting results that are not easily reconciled with plausible changes in pricing fundamentals. Several studies using U.S. data through the 1980s provide evidence that loss ratios and reported underwriting profit margins (e.g., one minus the combined ratio) exhibited second-order autoregression, implying a cyclical period of about six years (Venezian, 1985, Cummins and Outreville, 1987, Doherty and Kang, 1988). Analysis also suggests cyclical underwriting results in other countries (Cummins and Outreville, 1987, Leng and Meier, 2006, Meier, 2006a and 2006b, Meier and Outreville, 2006). Other studies suggest that underwriting results remain cyclical after controlling for the expected effects of changes in interest rates (Fields and Venezian, 1989, Smith, 1989; also see Winter, 1991a). Cummins and Outreville (1987) show conditions under which accounting and regulatory lags could generate a cycle in reported underwriting margins and present suggestive empirical results. Doherty and Kang (1988) argue that cyclical patterns in underwriting results reflect slow but presumably rational adjustment of premiums to changes in expected claim costs and interest rates.

**Unit Roots and Cointegration Analyses.** Early time series studies paid little attention to stationarity assumptions. Later work more carefully considers whether underwriting margins are stationary and, if not, whether they are cointegrated with macroeconomic factors. Assuming non-stationarity, Haley (1993, 1995) presents evidence that underwriting profit margins are negatively cointegrated with interest rates in the long run. Results for error correction models indicate a short-run relation between interest rates and underwriting margins. Grace and Hotchkiss (1995) provide evidence of cointegration between quarterly combined ratios and short-term interest rates, the consumer price index, and real GDP. Choi, Hardigree, and Thistle (2002) provide evidence that
underwriting profit margins are cointegrated with annual Treasury bond yields, but not with the ratio of capital to assets.

Harrington and Yu (2003) conduct extensive unit root tests of the series typically analyzed in underwriting cycle research, presenting evidence largely consistent with stationarity (absence of a unit root). In contrast, Leng (2006) presents evidence that combined ratios are non-stationary and subject to structural breaks. Jawad, Bruneau, and Sghaier (2009) provide evidence that premiums are cointegrated with interest rates using non-linear cointegration techniques. Lazar and Denuit (2012) analyze the dynamic relationship between U.S. property-casualty premiums, losses, GDP, and interest rates using both single equation and vector cointegration analyses. The results suggest long-term equilibrium between premiums, losses, and the general economy, and that premiums adjust quickly to long-term equilibrium.

**Capital Shocks and Capacity Constraints.** Economic models of premium volatility have generally focused on the potential effects of industry-wide shocks to capital. These capital shock models basically assume that industry supply depends on the amount of insurer capital and that industry supply is upward sloping in the short run because the stock of capital is costly to increase due to the costs of raising new capital. These features imply that shocks to capital (e.g., from catastrophes or unexpected changes in liability claim costs) affect the price and quantity of insurance supplied in the short run. Holding industry demand fixed, a backward shift in the supply curve due to a capital shock causes price to increase and quantity to decrease, which roughly describes “hard” insurance markets. Soft markets – low prices and high availability – either are not addressed by these models or are explained by periods of excess capital that is not paid out to shareholders because of capital exit costs.


---

43 Note that it is not clear why insurance underwriting margins would be non-stationary apart from structural breaks. A unit root would imply that shocks to ratios of insurance losses and expenses to premiums or GDP would be permanent after controlling for trend, and there is no obvious reason that shocks would be permanent.

44 All of the capital shock models are built on the assumption that external capital is costlier than internal capital, often justified with the logic of Myers and Majluf (1984). Capital shock models were motivated in significant part by the hard market in U.S. general liability insurance in the mid-1980s. A related line of work considers the effects of subsequent tort liability reforms (e.g., Born and Viscusi, 1994).
idea that insolvency risk depends on the amount of insurer capital because of uncertainty in claim costs and/or investment returns. Some models explicitly or implicitly assume that in the long run insurers choose capital levels to equate the marginal costs and benefits of additional capital.45 Others assume that insurer capital must satisfy a regulatory constraint on the probability of insolvency. Most models are static, with dynamic aspects of the market explained by periodic exogenous shocks. An exception is Winter (1994), which models the dynamics of the insurance market in a discrete time equilibrium model.

In empirical tests, Winter (1994) regresses an “economic loss ratio” (which discounts reported losses) on interest rates and the lagged values of insurer capital relative to its previous five-year average. Consistent with the prediction that higher prices (lower expected loss ratios) occur when capital is low, the coefficients on the lagged capital variables are positive and statistically significant in most of his specifications.46 Gron (1994b) obtains results with similar implications with different underwriting performance measures. Gron (1994a) examines aggregate time series data on underwriting profit margins for auto physical damage, auto liability, homeowners’ multiple peril, and other liability insurance. The evidence suggests that margins are negatively related to deviations of relative capacity (capital to GDP) from its normal level, which again is consistent with the notion that prices increase when capacity (insurer capital) is reduced.

Other capital shock models consider the potential for heterogeneous responses among firms. Doherty and Garven (1995) consider the effects of interest rate changes, which can influence capital by changing the value of assets and liabilities. They predict that interest rate changes will cause firm-specific capital shocks, as well as alter the long run equilibrium price of insurance. They use insurer panel data to estimate the sensitivity of insurer underwriting returns to interest rate changes. The results provide some evidence that return sensitivity is related to surplus duration.

45 Cagle and Harrington (1995) examine the extent to which the cost of a capital shock may be passed on to consumers in the form of higher prices. In their model, insurers choose an optimal level of capital based on the cost of holding capital and the benefits of protecting franchise value. They derive comparative statics for the upper bound effect on price of a shock to capital, assuming that demand is perfectly inelastic and that additional capital cannot be raised, and show that prices will not increase enough to fully offset shocks.

46 During the 1980s, however, the correlation between insurer capital and the economic loss ratio was negative. Winter argues that the 1980s can be explained in part by the omission of reinsurance capacity in his capital variables, a factor which also may have influenced the results of Cummins and Danzon (1997, see below).
Cummins and Danzon (1997) consider an insurer that enters a period with existing liabilities and a stock of capital and chooses the amount of new capital to raise and the price for new policies. Demand for coverage depends both on price and quality (insolvency risk). The benefit of additional capital is increased demand for new policies. The cost of additional capital is that the old policyholders (existing liabilities) experience reduced insolvency risk, while not paying additional premiums. Among other predictions, Cummins and Danzon show that prices could fall if a capital shock increases insolvency risk. They estimate a two-equation system using insurer level data, where price depends on lagged capital and additions to capital depend on the change in price. Their results indicate that insurers with more capital charge higher prices and that price is inversely related to deviations of capital from normal levels.

Froot and O’Connell (1997) test the extent to which shocks in one insurance market influence pricing in other markets. They present evidence that catastrophe reinsurance prices change across the board in response to shocks caused by specific types of catastrophes (e.g., a hurricane) or by catastrophes in specific regions (also see the discussion in Froot, 2001). Weiss and Chung (2004) provide evidence that reinsurance prices are negatively related to worldwide relative capacity.

**Price Cutting in Soft Markets.** The traditional view of underwriting cycles by insurance industry analysts is that supply expands when expectations of profits are favorable, that competition then drives prices down to the point where underwriting losses deplete capital, and that supply ultimately contracts in response to unfavorable profit expectations or to avert financial collapse. Price increases then replenish capital until price-cutting ensues again. The traditional explanation of supply contractions is largely consistent with shock models. The principle puzzle is why competition in soft markets would cause insurers to cut prices to the point where premiums and anticipated investment income are insufficient to finance optimal forecasts of claim costs and ensure a low probability of default.

Some authors suggest that a tendency towards inadequate prices might arise from differences in insurer expectations concerning the magnitude of future loss costs, from differences in insurer expectations concerning the magnitude of future loss costs, from differences in insurer expectations concerning the magnitude of future loss costs, from differences in insurer expectations concerning the magnitude of future loss costs, from differences in insurer expectations concerning the magnitude of future loss costs.

---

47 Gron and Winton (2001) provide a related analysis of the resulting risk overhang and market behavior.

48 Similarly, popular explanations of “cash flow underwriting” usually imply that insurers reduce rates too much in response to increases in interest rates. Winter’s model implies that hard markets that follow large shocks tend to be preceded by periods of excess capacity and soft prices. However, capital shocks should be unpredictable. Neither Winter’s model nor other shock stories can readily explain any second-order autoregression in underwriting results.
solvency incentives, or both. Harrington and Danzon (1994) develop and test hypotheses based on this intuition and the literature on optimal bidding and moral hazard within the framework of alleged underpricing of U.S. commercial general liability insurance during the early 1980s. In their analysis, some firms may price below cost because of moral hazard, which results from limited liability and risk-insensitive guaranty programs. Others may price below cost due to heterogeneous information concerning future claim costs, which results in low loss forecasts relative to optimal forecasts accompanied by winners’ curse effects. In response to underpricing by some firms, other firms may cut prices to preserve market share and thus avoid loss of quasi-rents from investments in tangible and intangible capital. Harrington and Danzon use cross-firm data from the early 1980s on premiums growth and loss forecast revisions for accidents in a given year to test whether moral hazard and/or heterogeneous information contributed to differences in general liability insurance prices and premium growth rates among firms. Their results provide some evidence consistent with their moral hazard hypothesis.

**Price Regulation**

Insurance markets generally are subject to substantial government regulation and supervision encompassing licensing of insurers and agents and brokers, solvency and capital standards, rates and policy forms (contract language), sales and claim practices, and requirements to issue coverage. A large literature has considered the effects of price regulation in insurance markets, especially in U.S. property/casualty insurance, where rates for many lines of insurance in many states are subject to regulatory prior approval; whereas other states allow rates to be used without prior regulatory approval. Most states had prior approval regulation during the 1950s and 1960s, when rate regulation often encouraged insurers to use rates developed collectively by rating bureaus within the framework of the limited exemption from U.S. antitrust law for the “business of insurance” (Joskow, 1973). A trend towards deregulation

---

49 Winter (1988, 1991a) mentions the possibility of heterogeneous information and winner's curse effects.
50 Harrington (2004c) presents evidence that higher firm-level premium growth for general liability insurance during the 1992-2001 soft market was reliably associated with higher loss forecast revisions, as would be expected if low-priced firms captured market share and ultimately experienced relatively high reported losses. Harrington, Danzon, and Epstein (2008) provide similar evidence using firm-level data for the U.S. medical malpractice insurance soft market during 1994-1999.
51 See Klein (2012, 2013) for overviews in the context of current issues.
began in the late 1960s and continued, albeit sporadically through the 1990s (see Harrington, 2000).

Rate regulation can affect an insurer’s average rate level in a given period. It also can affect rate differentials between groups of consumers by imposing limits on rates for particular groups through residual market mechanisms and/or by restricting risk classification.\footnote{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.} Research on rate regulation in the United States has generally focused on politically sensitive lines of business, in particular automobile and workers’ compensation insurance, and more recently homeowners’ insurance in catastrophe-prone regions. These lines have been subject to more restrictive rate regulation in many states, and they periodically have experienced large residual markets in some states, in significant part due to restrictive rate regulation.

Many studies compare loss ratios, most often for automobile insurance, in states with and without prior approval regulation to provide evidence of whether prior approval affects average rate levels in relation to claim costs.\footnote{Cummins (2002) and the studies therein consider this issue and other effects of rate regulation.} Hypotheses considered include that regulation raises rates due to capture by industry, that regulation persistently reduces rates due to consumer pressure, or that on average rate regulation will have little effect in competitively structured insurance markets. Some analyses indicate that regulation in some states and periods resulted in higher auto insurance loss ratios in states with prior approval rate regulation (e.g., during the mid-to-late 1970s and early 1980s; Harrington, 1987; Grabowski, Viscusi, and Evans, 1989) and that a significant number of insurers exited the auto insurance market in such states.

On the other hand, and consistent with an inherent inability of prior approval regulation to reduce insurance rates persistently without harmful effects on availability and quality, studies generally have found no persistent difference over time between average loss ratios in states with and without prior approval regulation. Harrington (2002), for example, analyzes automobile insurance loss ratios, residual market shares, and volatility of premium growth rates by type of rate regulation with annual state-level data during 1972-1998. Conditioning on several factors that could affect loss ratios, the average loss ratio in prior approval states was only slightly larger than in other states, with the difference primarily attributable to the 1970s and at most weakly significant in a statistical sense.
Other studies document larger residual markets in automobile insurance in states with prior approval regulation (Ippolito, 1979; Grabowski, Viscusi, and Evans, 1989; Harrington, 2002), with some degree of cross-subsidy flowing from voluntary markets to higher risk buyers in residual markets. Harrington and Danzon (2000b) and Danzon and Harrington (2001) consider the incentive effects of regulatory induced cross-subsidies to higher risk employers in workers’ compensation insurance residual markets. They provide evidence that higher subsidies are associated with greater growth in loss costs and increases in the proportion of employers that self-insure. Using annual, state-level panel data during 1980-1998, Weiss, Tennyson, and Regan (2010) provide evidence that rate regulation that suppresses premiums is associated with higher average loss costs and claim frequency for automobile insurance. Comparing Massachusetts and other states and towns within Massachusetts, Derrig and Tennyson (2011) provide evidence that regulatory-induced cross-subsidies to high risk drivers are likewise associated with higher automobile insurance loss costs.

Another avenue of inquiry is whether delays in the rate approval process under prior approval rate regulation could influence or even cause cyclical fluctuations in underwriting results (Cummins and Outreville, 1987). A number of studies analyze whether rate regulation affects cyclical movements in statewide loss ratios (e.g., Outreville, 1990; Tennyson, 1993). Such studies generally obtain mixed results concerning whether regulatory lag amplifies volatility in underwriting results.

**CAPITAL ADEQUACY AND REGULATION**

*Optimal Capital Decisions.* The theory of insurers’ optimal capital decisions focuses on the tradeoff between the benefits and cost of holding additional capital. The primary benefits include higher prices from risk sensitive policyholders and/or reducing the likelihood of financial distress, including the potential loss of franchise value. The primary costs include tax and agency costs. Building on the work of Borch (e.g., Borch, 1982; also see DeFinetti, 1957), Munch and Smallwood (1982; also see Munch and Smallwood, 1980) and Finsinger and Pauly (1984) model insurer capital decisions in the presence of costly financial distress assuming that insurers maximize value to shareholders, that policyholder demand is inelastic with respect to default risk, that investing financial capital to support insurance operations is costly, and that firms cannot add capital after claims are realized. In the event of insolvency, Munch and Smallwood (1982) assume a loss of goodwill. Finsinger and Pauly (1984) assume loss of an entry cost that
otherwise would allow the firm to continue operating (also see Tapiero, Zuckerman, and Kahane, 1978). In both models firms will commit capital ex ante to reduce the likelihood of insolvency.55

Doherty and Tinic (1981), Doherty (2000), and Tapiero, Kahane, and Jacques (1986) consider insurers’ capital decisions when demand for coverage is sensitive to insolvency risk. Garven (1987) analyzes insolvency risk within an agency cost framework in which shareholders, managers, sales personnel, and policyholders have different incentives regarding insolvency risk. Cagle and Harrington (1995) consider a static tradeoff model of capital decisions when insurers are exposed to loss of franchise value for the cases of risk sensitive and risk insensitive policyholder demand. Froot (2007) models capital decisions with risk sensitive demand, tax costs of holding capital, and adjustment costs if capital is raised following losses.

**Market Discipline.** The scope of market discipline provided by policyholders and capital markets is a key element affecting capital decisions in many models. Factors affecting overall market discipline include (1) potential loss of franchise value, which arises from insurers’ upfront investments in infrastructure and distribution, underwriting expertise and information, and reputation; (2) the risk sensitivity of demand, which depends on the ability of policyholders or their representatives to assess insolvency risk, the scope of any government guarantees of insurers’ obligations, and the scope of the judgment proof problem, where some buyers may rationally seek low prices regardless of insolvency risk; (3) the risk sensitivity of insurance intermediaries (agents and brokers), who are exposed to increased costs or reduced revenues from insurer financial distress; and (4) risk sensitive debt-holders, which provide debt finance primarily at the insurance holding company level (Harrington, 2004b, also see Eling, 2012).

A variety of studies provide empirical evidence on market discipline. Consistent with risk sensitive demand, studies that employ measures of property/casualty insurers’ premiums in relation to realized claim costs as proxies for the price of coverage have found that prices are negatively to insolvency risk (Sommer, 1996; Cummins, Phillips, and Allen, 1998; Cummins, Phillips, and Linn, 2006). Zanjani (2002a) provides evidence that life insurance policy termination rates are greater for insurers with lower A.M. Best financial strength ratings, although terminations are not related to rating changes. Epermanis and Harrington (2006) estimate abnormal premium growth surrounding changes in A.M. Best financial strength ratings for a large panel of property/casualty insurers during 1992-1999. They report evidence of

---

55 The literature on capital decisions by banks contains similar results (e.g., Herring and Vankudre, 1987).
significant premium declines in the year of and the year following rating downgrades. Consistent with greater risk sensitivity of demand for customers with more sophistication and less guaranty fund protection, premium declines were concentrated among commercial insurance. Premium declines were greater for firms with low pre-downgrade ratings, and especially pronounced for firms falling below an A– rating, a traditional industry benchmark for many corporate buyers. Eling and Schmit (2012) conduct similar tests for German insurers and obtain qualitatively similar results, although the responses to rating changes appear smaller in magnitude than in the United States.

The extent to which partial state government guarantees of U.S. insurers’ obligations dull policyholders’ incentives and/or induce excessive risk taking from moral hazard has been explored in a number of studies that exploit cross-state variation in insurance guaranty fund characteristics or adoption dates to test whether guarantees have increased risk taking. Lee, Mayers, and Smith (1997) provide evidence that asset risk increased for stock property/casualty insurers following the introduction of guaranty funds (also see Lee and Smith, 1999, and Downs and Sommer, 1999). Brewer, Mondschean, and Strahan (1997) provide evidence that life insurer asset risk is greater in states where guaranty fund assessments against surviving insurers are offset against state premium taxes and thus borne by taxpayers, which may reduce financially strong insurers’ incentives to press for efficient regulatory monitoring.56

Many studies and numerous anecdotes for particular insolvencies document unusually large premium growth prior to the insolvencies of some property/casualty insurers (e.g., A.M. Best, 1991; Bohn and Hall, 1999; also see Harrington and Danzon, 1994), which could in some cases be plausibly related to underpricing and excessive risk taking.57 The 1990-1991 insolvencies of First Executive Corporation, First Capital Corporation, and Mutual Benefit Life in the United States suggest some degree of ex ante excessive risk taking, but the evidence is not sharp. Mutual Benefit ended up meeting virtually all of its obligations. Both First Executive and First Capital might have remained viable without regulatory intervention despite the temporary collapse of the junk bond market, and their experience demonstrated that demand was risk sensitive: as more bad news surfaced, more policyholders withdrew their funds (DeAngelo, DeAngelo, and Gilson,

56 A few papers have analyzed the design of state guaranty funds, including their ex post assessment feature (see, e.g., Bernier and Mahfoudhi, 2010).

57 On the other hand, Epermanis and Harrington (2006) found no evidence of excessive risk taking in the form of rapid commercial or personal lines premium growth following downgrades of A– or low-rated insurers.
Consistent with market discipline by equity holders, Fenn and Cole (1994) and Brewer and Jackson (2002) provide evidence that life insurer stock price declines during the high-yield bond and commercial real estate market slumps of 1989-1991 were concentrated among firms with problem assets.

**Solvency Regulation and Capital Requirements.** There are three main strands of literature on insurance company insolvencies and solvency regulation: (1) analyses (generally descriptive) of the causes of insolvencies; (2) insolvency prediction models, including those that incorporate information from regulatory solvency monitoring systems and risk-based capital requirements; and (3) the design and performance of regulatory risk-based capital (RBC) requirements.


Cummins, Harrington, and Niehaus (1994) provide an early conceptual evaluation of the NAIC’s RBC requirements. More recent work contrasts the NAIC RBC standards and capital standards in the E.U.’s forthcoming Solvency II regime (and in some cases other countries; Eling, Schmeiser, and Schmit, 2007; Eling and Holzmuller, 2008; Holzmuller, 2009; Cummins and Phillips, 2009). The NAIC system is a deterministic and formulaic approach that combines fixed weights on a variety of risk measures based on statutory accounting to obtain a required RBC.
amount. The Solvency II regime, patterned to a great extent after the Basel II (and III) regime for banks, determines required capital using market-based valuation and Value-at-Risk (VaR) models, and it encourages firms to use internal capital models. Cummins and Phillips (2009, also see Holzmuller, 2009) recommend adoption of market valuation, stochastic risk modeling, and internal capital models in the United States.\(^{58}\)

Harrington (2004a) considers the optimal stringency of capital standards in relation to the degree of market discipline when imperfect information concerning capital adequacy results in costly Type 1 and 2 errors from any system of capital standards (i.e., failing to constrain some inadequately capitalized insurers and inefficiently distorting the decisions of some adequately capitalized insurers). His model predicts that cost minimizing capital standards will be less stringent the greater is the proportion of insurers that would be adequately capitalized without regulation. An implication is that optimal capital standards should not bind most insurers in a market characterized by strong market discipline. He therefore argues that U.S. RBC system, where most insurers hold significantly more capital that the required minimums (Harrington, 2004a; Cummins and Phillips, 2009), is consistent with strong market discipline in U.S. insurance markets.\(^{59}\)

**Systemic Risk.** The 2007-2009 financial crisis in general and the 2008 collapse of American International Group (AIG) in particular stimulated substantial research and analysis of the extent to which insurance involves systemic risk. Those events also led to the adoption in the United States of regulation requiring the identification of systemically important financial institutions (“SIFIs”), including insurance entities, and heightened regulatory standards for such entities.\(^{60}\)

Although there is no uniform definition, the term “systemic risk” generally is used broadly to encompass the risk of any large, macroeconomic shock that affects financial stability and the risk arising from extensive interdependencies or “interconnectedness” among firms, with an attendant risk of contagion and significant economic spillovers on the real economy. Cummins and Weiss

---

\(^{58}\) Klein (2012) discusses insurance regulatory modernization initiatives in the United States in the context of challenges posed by Solvency II.

\(^{59}\) De Haan and Kakes (2010) document that the vast majority of Dutch insurers hold much more capital than required by the E.U.'s pre-Solvency II rules.

\(^{60}\) Similarly, the International Association of Insurance Supervisors and the Financial Stability Board are in the process of identifying global systemically important insurers. The systemic risk issue for insurance received some attention prior the financial crisis. Harrington (2004a), for example, contrasts systemic risk among property/casualty insurers, life insurers, reinsurers, and banks.
(2013b), for example, define systemic risk as “the risk that an event will trigger a loss of economic value or confidence in a substantial segment of the financial system that is serious enough to have significant adverse effects on the real economy with a high probability.” They suggest that primary indicators of systemic risk at the firm level include size (volume of exposures), interconnectedness, and a lack of substitutability for a firm’s services.61

There is a distinction between the risk of common shocks to the economy, such as widespread reductions in housing prices or large changes in interest rates or foreign exchange, which have the potential to directly harm large numbers of people and firms, and financial risk that arises from interconnectedness and contagion.62 However, it often is difficult to sort out any contagion effects from the effects of common shocks, and broad definitions of systemic risk encompass both sources of risk. Harrington (2009) discusses uncertainty about whether AIG’s credit default swaps and securities lending presented significant risk of contagion and the extent to which an AIG bankruptcy would have had significant adverse effects beyond its counterparties, or the extent to which its counterparties had hedged or otherwise reduced their exposure to AIG.63

Analyses generally conclude that the core activities of insurers pose little systemic risk, especially compared with banking, in part because many insurers hold relatively large amounts of capital in relation to their liabilities and have relatively little exposure to short-term liabilities, reducing their vulnerability to shocks (Swiss Re, 2003; Harrington, 2004a; Geneva Association, 2010, 2012; Grace, 2010; Cummins and Weiss, 2013b). Based on a detailed review and analysis, for example, Cummins and Weiss (2013b) conclude that “the core activities of U.S. insurers do not

61 Section 113 of the 2010 U.S. Dodd-Frank Act authorizes the newly created Financial Stability Oversight Council (FSOC) to designate a nonbank financial company as systemically significant and subject to enhanced supervision by the Federal Reserve if “material financial distress at the nonbank financial company” or “the nature, scope, size, scale, concentration, interconnectedness, or mix of the activities of the nonbank financial company could pose a threat to the financial stability of the United States.” The FSOC indentified six broad risk categories for determining systemic importance: size, lack of substitutes for the firm’s services and products, interconnectedness with other financial firms, leverage, liquidity risk and maturity mismatch, and existing regulatory scrutiny.

62 There are at least four sources of potential contagion that could contribute to systemic risk (e.g., Kaufman, 1994), including: (1) asset price contagion (“fire sales”); (2) counterparty contagion; (3) information-based contagion (the revelation of financial problems at some institutions creates uncertainty about the effects on counterparties); and (4) irrational contagion (investors and/or customers withdraw funds without regard to risk). Ellul, Jotikashthira, and Lundblad (2011) provide evidence of regulatory induced fire sales of downgraded corporate bonds by life insurers.

63 Without intervention by the U.S. government, many more of AIG’s insurance customers might have terminated or declined to renew their policies, but that by itself would not imply contagion, or that those customers would be significantly harmed.
pose systemic risk.” They also conclude, however, that “life insurers are vulnerable to intra-sector crises” and that “both property-casualty and life insurers are vulnerable to reinsurance crises arising from counterparty credit risk.”

Research on financial institutions’ stock prices provides evidence of interconnectedness among insurers and other financial firms and develops new metrics for measuring systemic risk with stock price data. Billio, et al (2011), for example, use principal components analysis and Granger causality tests to analyze stock returns for insurers, banks, securities brokers, and hedge funds during 1994-2008. They find evidence of causal relationships between the sectors during 2001-2008 but not 1994-2000, and they identify several insurers as systemically important. Acharya, et al (2010) develop a measure of systemic risk (systemic or marginal expected shortfall) to reflect a firm’s tendency to lose value when the overall market suffers large losses. Their analysis of stock returns for insurers and other financial firms during 2006-2008 suggests that insurance firms were the least systemically risky. Insurers with the largest systemic risk measures had significant activity in credit derivatives and financial guarantees. Cummins and Weiss (2013b) conclude that the Billio, et al and Acharya, et al studies strongly suggest that insurance firms can be a source of systemic risk from non-core activities and that interconnectedness among financial firms goes beyond exposure to common shocks.

**Securitization and Insurance-Linked Securities**

Innovations in hybrid reinsurance-financial products and insurance-linked securities (ILS) represent a major development in insurance and reinsurance markets during the past two decades. Hybrid products and structures include finite risk reinsurance (which smooth financial results over multiple years with limited risk transfer), retrospective excess of loss coverage (protection against adverse loss development after events have occurred), loss portfolio transfers (transfers of entire blocks of claim liabilities), multiple trigger products (protection, for example, against the simultaneous occurrence of an insurance event and a financial market event), industry loss warranties (multiple trigger products combing industry and indemnity triggers), and side cars (reinsurance companies established by institutional investors to provide coverage to a single insurer or reinsurer sponsor). ILS products include contingent capital arrangements (which allow an insurer to raise capital at pre-determined terms following a specified event), catastrophe

---

64 Chen, et al (2012) use Granger causality tests to examine interconnectedness between banks and insurers using data on credit default swap spreads.
futures and options (providing payoffs dependent on a specified index or trigger), and catastrophe bonds (bondholders receive less than full repayment of principal and/or interest if specified catastrophe losses occur).

Many of these innovations offer the potential for more efficient risk transfer through reduced moral hazard and adverse selection, improved diversification, reduced frictional costs of risk transfer, and/or enhanced capital market opportunities for investors. Cummins and Barrieu (2013, also see Cummins and Weiss, 2009, and Cummins and Trainar, 2009) provide a detailed review of hybrid and ILS product features, markets, and research. This section provides an overview of selected research on catastrophe futures/options and catastrophe (cat) bonds.

**Catastrophe Futures, Options, and Basis Risk.** Early research on catastrophe derivatives coincided with Hurricane Andrew and the introduction by the Chicago Board of Trade of catastrophe futures contracts in 1992 (replaced by catastrophe options in 1994, which in turn were withdrawn in 2000). D’Arcy and France (1992), Cox and Schwebach (1992), and Niehaus and Mann (1992) identify potential benefits of insurance derivatives compared with reinsurance, including possibly lower transaction costs and reduced counterparty risk, and they discuss the tradeoff between moral hazard and basis risk, possible regulatory barriers, and potential hedging strategies. Harrington, Mann, and Niehaus (1995) consider the potential benefits of catastrophe futures and options compared with reinsurance or holding additional equity capital, and they provide evidence of potential hedging effectiveness by analyzing national loss ratios for alternative lines of business.

Harrington and Niehaus (1999) analyze the relationship between U.S. insurers’ loss ratios, state-level catastrophe losses, and industry loss ratios in different geographic areas. The results suggest that basis risk with state-specific catastrophe derivatives would likely be manageable for many insurers, in particular for homeowners’ insurance. Cummins, Lalonde, and Phillips (2004) analyze basis risk for catastrophe loss indexes for insurers writing windstorm insurance in Florida using exposure level data and simulated losses from a catastrophe modeling firm. The results suggest significant basis risk for smaller insurers.

Doherty (1997, 2000) emphasizes differences in basis risk associated with different types of triggers in ILS contracts. Contracts with indemnity triggers, which link payoffs to a specific insurer or reinsurer’s losses, minimize basis risk, but they aggravate moral hazard. Contracts with non-indemnity triggers (e.g., tied to an industry index) reduce moral hazard but increase basis.
risk. Doherty and Richter (2002) consider whether optimal hedging could involve a combination of indemnity and industry-based triggers.

**Cat Bonds.** While catastrophe futures and options have failed to gain market traction, the market for cat bonds appears viable and could be poised for additional growth. The basic structure of a cat bond involves the issuer (e.g., a reinsurer) issuing bonds through a fully collateralized special purpose vehicle, with the bonds providing the issuer an embedded call option that is triggered by a defined catastrophe event (see Cummins and Barrieu, 2013, for details). A variety of different triggers are used, including industry loss triggers, indemnity triggers, and parametric triggers based on physical severity levels. As a result of the embedded option, full repayment of principal and/or interest is contingent on whether the defined event occurs during the life of the bond. The resulting transfer of catastrophe risk to the bondholders provides investors with a “pure play” in the underlying risk, and it can achieve the tax advantages of debt vs. equity finance. Cat bonds with non-indemnity triggers also might reduce the adverse effects of asymmetric information (Finken and Lau, 2009). On the other hand, Lakdawalla and Zanjani (2006) explain that full collateralization of cat bonds “abandons the reinsurance principle of economizing on collateral through diversification of risk transfer.”

Research on the magnitude of cat bonds spreads over LIBOR and comparisons to catastrophe reinsurance prices identify two puzzles (Froot and O’Connell, 2008; Cummins and Mahul, 2008; Cummins and Barrieu, 2013). First, cat bond spreads are very high on average compared to those that might be expected given that cat bond returns should have little correlation with market returns. Second, cat bond spreads appear to mimic reinsurance rate changes in soft and hard markets. Dieckmann (2008) analyzes a habit and consumption-based asset pricing model in which cat bond losses tend to occur when catastrophe losses reduce consumption, thus increasing cat bond risk and possibly helping to explain large spreads even though catastrophe losses are small relative to aggregate consumption.

**INSURER EFFICIENCY, DISTRIBUTION, ORGANIZATIONAL FORM, AND CORPORATE GOVERNANCE**

**Efficiency Studies.** Joskow’s (1973) seminal analysis of the U.S. property/casualty 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. Joskow estimated simple models of insurer
operating expense ratios, concluding that the industry was characterized by constant returns to scale. Following Joskow, other early studies estimated insurers’ cost functions with cross-sectional accounting data (e.g., Doherty, 1981; Johnson, Flanigan, and Weisbart, 1981). Analysis of insurers’ operating efficiency then shifted to more elaborate models estimated with stochastic frontier analysis (SFA) and/or data envelope analysis (DEA).

Cummins and Weiss (2013a) review insurance efficiency studies in property/casualty and life insurance in nearly 20 countries. They identify 53 studies during the period 2000-2011, in addition to 21 studies considered earlier by Cummins and Weiss (2000). A majority of the 74 studies employ DEA as the primary methodology. Thirty-one studies were published during 2008-2011. Twenty studies focus on the effects of organizational form or corporate governance on estimated efficiency; 12 focus on the general level of efficiency over time; 12 focus on market structure, and 5 focus on economies of scale and scope. About two-thirds of the studies use loss, benefit, or claim-based measures of output, with 11 using premiums. The number and diversity of insurance efficiency studies obviates any high-level summary. Some specific studies dealing with distribution and organizational form are noted below.

**Insurance Distribution.** Insurance markets exhibit a wide variety of distribution methods across lines of business and countries. Hilliard, Regan, and Tennyson (2013) review the distribution literature, including studies of the economic rationales for different distributions systems, insurer relationships with and compensation of agents and brokers, and regulation of distribution. They also provide background information on the prevalence of alternative distribution methods in different lines of business in the Unites States and other countries. The literature generally focuses on differences between direct writing systems, where insurers primarily rely on exclusive agents, employee sales forces, or telephone/mail/internet systems, versus independent agency and brokerage systems, where intermediaries have relationships with multiple insurers.

Numerous studies compare insurers’ costs for different distribution methods, especially for direct writing versus independent agents and brokers in U.S. property/casualty insurance. Joskow (1973) provided early evidence that expense ratios were lower for direct writers than for independent agency insurers. Cummins and VanDerhei (1979) estimate more elaborate models of insurers’ expenses using pooled cross-section and time-series data. The results again indicate significantly lower expense ratios for direct writers. Later analyses by Barresse and Nelson
(1992) and Regan (1999) obtain similar results with richer measures of distribution type and data for more lines of business.

Joskow (1973) argued that differences in operating costs between direct writers and independent agency insurers could not be explained by differences in service. 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. He argued 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).

Cummins and VanDerheij (1979) assume that lower operating expenses for direct writers provide prima facie evidence of superior efficiency. On the other hand, based on stochastic frontier analysis of cost and profit efficiency for alternative distribution systems, Berger, Cummins, and Weiss (1997) provide evidence that while insurers with independent agents and brokers were less cost efficient than direct writers, there was no difference in profit efficiency. They interpret their overall results as consistent with higher costs being attributable to higher quality for independent agency and brokerage insurers.

Other work considers whether higher costs of independent agency and brokerage systems are associated with specific dimensions of service quality. Etgar (1976) provides evidence that independent agents are significantly more likely to assist policyholders in claims settlement than exclusive agents. Barrese, Doerpinghaus, and Nelson (1995) present evidence using data on consumer complaints to regulators in five states that insurers using independent agents receive fewer complaints. More recently, Eckardt and Rathke-Doppner (2010) provide evidence that independent agents in Germany provide better service quality, while exclusive agents provide a broader array of services.

Regarding insurers’ choice of distribution systems and co-existence of multiple systems, Marvel (1982) posits that direct writing helps internalize the benefits to the firm of advertising and promotion, implying that direct writing will be more prevalent when such activities are more important in generating sales. Grossman and Hart (1986) predict that ownership of customer lists
by the insurer or intermediary will be allocated to the party for which ownership is most likely to promote investments that maximize value and that agents will be allocated ownership when agent services are a more important determinant of value than insurers’ investments in brand formation. Sass and Gisser (1989) consider how direct writing depends on sufficient density of customers to offer agents greater volume potential at lower commission rates, and they provide evidence that market size and customer density are positively related to direct writer market shares.

Regan and Tennyson (1996) argue that independent agency systems provide greater incentives for agents to exert effort in risk selection. The reason is that independent agents’ have the ability to capture more profits from such efforts in view of their ability to influence consumers’ choices among competing firms. They predict that independent agency systems enhance efficiency when they have unverifiable, subjective information that is important to underwriting. Using state level data on market shares of direct writers during 1980-1987, they provide evidence that direct writer market shares are lower in states where exposures are more heterogeneous and complex.

Kim, Mayers, and Smith (1996) posit that independent agents and brokers will be used when agent monitoring of insurers is important to consumers and that such monitoring is more important for stockholder owned than mutual insurers. Using U.S. insurer level data for 1981, they provide evidence that direct writing is positively related to mutual ownership. Regan (1997) further considers transactions cost explanations of distribution choice and provides evidence with insurer level data that direct writing is positively related to insurers’ investments in advertising and technology and mutual ownership. Regan and Tzeng (1999) consider endogenous choice of distribution and organizational form and provide evidence that product complexity is more closely related to distribution choice than organizational form.

Cummins and Doherty (2006) provide detailed analysis of the economics of distribution and the forms of commission-based compensation in the U.S. commercial property/casualty insurance. Their work was motivated by allegations in 2004 by then New York Attorney General Elliot Spitzer that certain insurers and brokers had used contingent commissions to enforce anticompetitive pricing and market allocation schemes. They stress the role of independent agents and brokers in matching policyholders with insurers that have capabilities of meeting their

---

65 Posey and Tennyson (1998) model the influence of search costs on distribution system choice, predicting that low production and search costs will be associated with direct writing. They provide suggestive empirical evidence.
needs. They explain how contingent commissions, which make compensation to agents and brokers contingent on the volume, profitability, and/or persistency of business placed with an insurer, can help align insurer and intermediary interests, in part by reducing informational asymmetries between insurers and buyers and associated adverse selection.

A few studies have begun to explore the use and effects of internet technology on insurance distribution and market outcomes. Brown and Goolsbee (2002) provide evidence that internet comparison shopping in the United States significantly reduced term life insurance prices. Using data on internet investment and organizational characteristics for a sample of insurers, Forman and Gron (2011) provide evidence that direct writing speeds adoption of consumer internet applications that complement existing distribution.

**Alternative Organizational Forms.** In addition to significant variation in distribution methods, insurance markets are characterized by a variety of organizational forms, including stockholder ownership, mutual ownership, reciprocals, and Lloyds associations. Mayers and Smith (2013) describe the main types and survey the literature on choice of organization form, with an emphasis on the ability of alternative forms to reduce contracting costs associated with potential incentive conflicts among owners (residual claimants), managers, and policyholders.

Mayers and Smith (1981) consider the ability of alternative forms of ownership to minimize the cost of incentive conflicts (also see Fama and Jensen, 1983). They predict that mutuals will specialize in lines of insurance where managers have limited discretion to pursue their own interests at the expense of owners, whereas Lloyds associations should have greater penetration in lines where managers have the most discretion. Mayers and Smith (1988) discuss further the management discretion hypothesis, and they develop and test hypotheses concerning product specialization and geographic concentration across ownership types (also see Mayers and Smith, 1986). They obtain some evidence consistent with the prediction, for example, that mutuals specialize in lines with low managerial discretion.

Mayers and Smith (1990) analyze reinsurance purchases for U.S. property/casualty insurers (also see Mayers and Smith, 1982), providing evidence that Lloyds organizations reinsure the most and

---

66 Hansmann (1985), on the other hand, provided detailed discussion of the possible role of mutual ownership in reducing conflicts between owners and policyholders over the level of insurer default risk. He also considered the possible ability of mutual ownership to facilitate risk selection during the formative years of U.S. insurance markets.
that stock insurers reinsure less than mutuals. Lamm-Tennant and Starks (1993) provide evidence that stockholder owned insurers write more business than mutuals in lines of business with greater underwriting risk. The authors interpret that result as consistent with the managerial discretion hypothesis (lower risk lines involve less managerial discretion). Kleffner and Doherty (1996) find that stock insurers write more earthquake insurance than mutuals, but they interpret the result as consistent with stock insurers having superior access to capital following potentially large losses. Harrington and Niehaus (2002) provide evidence using U.S. property/casualty insurance data that mutual insurers hold more capital compared with stock insurers and adjust capital more slowly to target levels, interpreting these findings to mutuals’ greater costs of obtaining capital after adverse loss experience. Zanjani (2007) discusses the role of mutual firms’ inferior access to capital in his historical analysis of organizational form in life insurance.

A variety of studies have analyzed within-firm changes in organizational form. Mayers and Smith (1986) analyze the effects of insurer conversion from stock to mutual ownership on returns to stockholders, premium volume and persistence, product mix, and management turnover, concluding on average that such conversions were efficiency enhancing. Mayers and Smith (2002) provide evidence that mutual-to-stock conversions of property/casualty insurers were in large part motivated by the desire to achieve greater access to capital. Viswanathan and Cummins (2003) provide evidence that capital and liquidity constraints contributed to U.S. mutual insurer conversions to stock ownership for both property/casualty and life/health insurers (also see McNamara and Rhee, 2002). Erhemjamts and Phillips (2009) provide evidence that U.S. life insurer mutual-to-stock conversions were likewise motivated by access to capital concerns.

Cummins, Weiss, and Zi (1999) use DEA methods to analyze the relative efficiency of stock and mutual insurers. They provide evidence that stock and mutual firms have distinct production and cost frontiers, consistent with their having distinct technologies. Erhemjamts and Leverty (2010) use DEA methods to provide evidence that greater estimated cost efficiency for U.S. stock life insurers than mutual life insurers during 1995-2004 contributed to mutual-to-stock conversions, along with access to capital concerns.

**Corporate Governance.** Corporate governance is a new research topic in the insurance literature. It was motivated by corporate scandals around the world and the recent financial crisis in the

---

financial markets. Resources expropriations by well-paid executives and their risk-taking behavior motivated many studies on corporate governance in financial markets.

N. Boubakri (2013) offers a survey of the literature on the nature of corporate governance in the insurance industry. This new subject was covered extensively in a special issue of the *Journal of Risk and Insurance* in 2011. The focus is on several corporate governance mechanisms such as the composition of the Board of Directors, CEO compensation, and ownership structure. The impact of such mechanisms on insurers’ performance and risk taking is also discussed. Several avenues of future research are identified.

**CONCLUSIONS**

The theoretical literature on insurance has advanced enormously since the early work of Arrow and Borch. Models of optimal insurance and insurance demand by individuals have become highly refined, including detailed analysis of the influences of transaction costs, moral hazard, asymmetric information, uninsurable background risk, state dependent utility, and other issues. The theory of insurance and resource allocation also exhibits considerable sophistication. Substantial progress has likewise been made on modeling the demand for insurance and hedging by corporations. As is typically the case, empirical work has in many instances lagged theoretical developments, in large part due to limitations associated with available data. The increased availability of micro data, however, has led to significant progress, and it bodes well for further empirical work.

While less mature than theoretical work on demand and market equilibrium, significant theoretical literatures also have evolved to increase our understanding of insurance supply, including pricing, capital decisions, regulation, efficiency, distribution, and organizational form. Significant progress has been made in empirical work to test hypotheses generated by theoretical work on supply-related issues. Increasing availability of data again bodes well for future empirical work to test extant theories and new theoretical developments. Additional innovations in theoretical and empirical work also can be expected on the growing markets for insurance-lined securities and their effects on primary insurance and reinsurance markets. The overall evolution of the broad literature on insurance and insurance markets suggests continued progress in achieving a fuller understanding of existing arrangements and future developments in the years ahead.
References

A.M. Best Company (1991), *Best’s Insolvency Study*, A.M. Best Co.


Borch, K. (1990), Economics of Insurance, Amsterdam: North Holland.


Binding?” *Journal of Banking and Finance* 34, 1618-1627.


