Developments in risk and insurance economics: The past 50 years*

Henri Loubergé
GSEM – Geneva School of Economics and Management

Georges Dionne
Canada Research Chair in Risk Management, Finance Department
HEC Montréal

31 January 2024

Abstract

The chapter reviews the evolution in risk and insurance economics over the past 50 years, first recalling the situation in 1973, then presenting the developments and new approaches that have flourished since then. We argue that these developments were only possible because steady advances were made in the economics of risk and uncertainty and in financial theory. Insurance economics has grown in importance to become a central theme in modern economics, providing not only practical examples and original data to illustrate new theories, but also inspiring new ideas that are relevant to the overall economy.

Keywords: Insurance economics, optimal insurance protection, optimal self-protection, insurance pricing, insurance demand, economics of risk and uncertainty, financial economics, risk management, asymmetric information, insurance markets, climate finance.

JEL classification numbers: A33, B15, D10, D20, D80, D82, G22, G32, G52, L22.

1. Introduction

In the early seventies, some 50 years ago, the economics of risk and insurance was still embryonic. Indeed, when the International Association for the Study of Insurance Economics (known as the “Geneva Association”) was founded in 1973, one of its promoters’ main goals was to foster the development of risk and insurance education in economics curricula. In particular, there was a clear need to develop the understanding of risk and insurance issues among the future partners of the insurance industry. It also seemed necessary to attract the attention of economists to risk and insurance as a stimulating and promising research field.

At that time, some attempts to link insurance to general economic theory had already been made, but they were still scarce. The books written by Pfeffer (1956), Mahr (1964), Greene (1971), and Carter (1972), or the one edited by Hammond (1968) tried to bridge the gap. Corporate risk management started to be considered seriously as a branch of study, at least in the United States. See Mehr and Hedges (1963) and Greene (1973) for early references. The main obstacle was obvious: traditional economic theory was based on the assumption of perfect knowledge—with some ad hoc departures from this assumption, as in the theory of imperfect competition or in Keynesian macroeconomics. In order for risk and insurance issues to be integrated into general economics, the theory of risk had to develop and gain a position at the heart of economic theory. The foundations were already at hand: the von Neumann-Morgenstern (1947) and Savage (1954) theory of behavior under uncertainty, the Friedman-Savage (1948) application to risk attitudes, Pratt’s (1964) analysis of risk aversion, Rothschild and Stiglitz’s (1970) characterization of increases in risk, and the Arrow (1953) and Debreu (1959) model of general equilibrium under uncertainty. These approaches had already started to bring about a first revolution in the study of finance, with the Markowitz (1959) model of portfolio selection and the Sharpe (1964)–Lintner (1965)–Mossin (1966) model of equilibrium capital asset pricing (the CAPM). With the benefit of hindsight, we now know that they
provided the starting point for the accomplishment of one of the Geneva Association’s long-term objectives: the integration of risk and insurance research into mainstream economic theory. The papers presented for the 50th anniversary of the European Group of Risk and Insurance Economists (EGRIE), founded by The Geneva Association, provide evidence of this achievement. See, in particular, Eling (2024), Gollier (2024), Peter (2024a), and Rothschild (2024).

The purpose of this chapter is to first remind the reader of the situation of insurance economics in 1973 (Section 2), and then to summarize its major developments since then in the three main areas of investigation that could be defined at that time: optimal insurance and protection (Section 3); market equilibrium under asymmetric information (Section 4); and insurance market structure (Section 5). A sixth section reviews financial economics by focusing on the new approaches that resulted from the growing integration of insurance and finance. The seventh section concludes by discussing recent developments in insurance markets. The discussion is mainly concentrated on risk and insurance economics issues as they relate to property-liability insurance. Due to limitations of space and time, two important related topics were omitted from this survey: health economics and social security. Additionally, life insurance is covered only partially.

2. Insurance economics in 1973

In 1973, the economic theory of insurance had already started to develop from six seminal papers: Borch (1962), Arrow (1963), Mossin (1968), Akerlof (1970), Ehrlich and Becker (1972), and Joskow (1973).¹ All these articles were based on the expected utility paradigm.

¹ Note that three of these seven authors, George Akerlof, Kenneth Arrow, and Gary Becker, later received the Nobel Prize for economics.
Following these articles, a whole series of important contributions were published, signaling that the development of an economic theory of risk and insurance was underway.

2.1. Borch (1962)

In his 1962 *Econometrica* article, “Equilibrium in Reinsurance Markets,” Karl Borch showed how Arrow’s (1953) model of general equilibrium under uncertainty could be applied to the problem of risk sharing among reinsurers. But generations of economists later learned that this insurance application had far-reaching implications for the general economy. In 1953, Arrow had shown that financial markets provide an efficient tool to reach a Pareto-optimal allocation of risks in the economy. Nine years later, Borch’s theorem demonstrated how the mechanism could be organized in the practice of insurance.

The main argument is the following. In a population of risk-averse agents, only social risks matter. Individual risks do not really matter because they can be diversified away using the insurance markets (the reinsurance pool of Borch’s paper). But social risks—those affecting the economy at large—cannot be diversified; they must be shared among participants to the pool. Borch’s theorem on Pareto-optimal risk exchanges implies that the sharing rule is based on participants’ risk tolerances (Wilson, 1968). Each individual (reinsurer) gets a share of the social risk (the reinsurance pool) in proportion to its absolute risk tolerance, which is the inverse of absolute risk aversion. If all individual utility functions belong to a certain class (later known as the HARA class and including the most widely used utility functions), the sharing rule is

---

2 See Gollier (1992) for a review of the economic theory of risk exchanges, Drèze (1979) for an application to human capital, and Drèze (1990) for an application to securities and labor markets.

3 Actually, Borch's theorem was already present in Borch (1960), but the 1960 article was primarily written for actuaries, whereas the 1962 *Econometrica* paper was addressed to economists.

4 Climate risk could be an example of a social risk that cannot be totally diversified. See Cummins et al. (2002) and Dionne and Desjardins (2022) for applications of the Borch model to climate risk management.

5 HARA = Hyperbolic Absolute Risk Aversion. As noted by Drèze (1990), the linearity of the sharing rule follows from the linearity of the absolute risk tolerance implied by hyperbolic absolute risk aversion.
linear. The CAPM, which for long was the dominant paradigm in finance theory, represents a special case of this general result.

Borch’s paper is the cornerstone of insurance economics. It can conveniently be used to show how the insurance mechanism of risk pooling is part of a more global financial mechanism of risk allocation and how a distinction can nevertheless be made between insurance institutions and other financial institutions. For this reason, it can also be used to clarify ideas on a hotly debated issue: the links between finance and insurance (see Section 6 below).

Prior to 1973, Borch’s seminal contribution found its main insurance economics extensions in the papers by Arrow (1970), Kihlstrom and Pauly (1971), and Marshall (1974). Arrow (1970) explicitly defined insurance contracts as conditional claims, i.e., an exchange of present money against conditional future money. Kihlstrom and Pauly (1971) introduced information costs in the risk-sharing model, arguing that economies of scale in the processing of information explain why insurance companies exist. In 1974, Marshall further extended this analysis by introducing a distinction between two modes of insurance operations: reserves and mutualization. Under the reserve mode, aggregate risk is transferred to external risk bearers (investors). With mutualization, external transfer does not apply, or cannot apply: aggregate losses are shared among insureds.

2.2. Arrow (1963)

The article published in 1963 by Kenneth Arrow in *The American Economic Review*, under the title “Uncertainty and the Welfare Economics of Medical Care,” represents the second point of departure for risk and insurance economics. This work may be credited with at least three contributions. Firstly, the article provided, for the first time, what has now become the

---

6 The question of whether or not “institutions” are needed to allocate risks in the economy was tackled later in the finance literature.

7 The applications of Borch’s theorem in the actuarial literature are reviewed by Lemaire (1990).
most famous result in the theory of insurance demand: if the insurance premium is loaded, using a fixed-percentage loading above the actuarial value of the policy, then it is optimal for an expected utility–maximizing insured to remain partially at risk, i.e., to purchase incomplete insurance coverage. More specifically, Arrow (1963) proved that full insurance coverage above a deductible is optimal in this case. Secondly, Arrow (1963) also proved that, when the insured and insurer are both risk-averse expected-utility maximizers, Borch’s theorem applies: the Pareto-optimal contract involves both a deductible and coinsurance of the risk above the deductible. This result was later extended by Moffet (1979) and Raviv (1979), and generalized by Gollier and Schlesinger (1996) and by Schlesinger (1997) under the less restrictive assumption of risk aversion.³ Thirdly, the Arrow article was also seminal in that it introduced asymmetric information into the picture. Arrow (1963) noted that transaction costs and risk aversion on the insurer’s side explained the incomplete risk transfer, but he also emphasized that moral hazard and adverse selection represented major obstacles to the smooth running of the insurance mechanism. By attracting the attention of economists to these problems, he paved the way for more focused work by Pauly (1968) and Spence and Zeckhauser (1971) on moral hazard, and by Akerlof (1970) on adverse selection.

2.3. Mossin (1968)

The paper by Jan Mossin, “Aspects of Rational Insurance Purchasing,” published in 1968 in The Journal of Political Economy, is generally considered the seminal paper on the theory of insurance demand, although some of Mossin’s results were also implicit in Arrow (1963) and explicit in another paper on insurance demand published earlier the same year in the same

³ More precisely, Schlesinger (1997) considers one version of Arrow’s theorem: the case where the insurer is risk neutral and the insured is risk averse (risk aversion being defined by Schlesinger as preferences consistent with second-degree stochastic dominance). In this case, a straight deductible policy is optimal whenever the insurer’s management costs are proportional to the indemnity payment.
Mossin’s paper is mainly famous for having shown that 1) partial insurance coverage is optimal for a risk-averse expected-utility maximizer when the insurance premium is such that a positive proportional loading applies to the actuarial value of the policy; and 2) insurance is an inferior good if the individual has decreasing absolute risk aversion (DARA). It was later pointed out (see below) that these strong results are respectively based on the implicit assumptions that the individual faces only one risk and that the amount of risk is fixed (unrelated to wealth or income).

2.4. Akerlof (1970)


His most famous example was a used car market where buyers cannot distinguish the different qualities of used cars, and all cars are sold at the same average price. The presence of bad used cars (lemons) can drive the good used cars out of the market when the difference in quality is high, because the average price under asymmetric information is too low for sellers of good used cars. This example explains the development of various institutions to counteract the reduction in average quality, such as guarantees and brand-name goods.

Akerlof’s main example in insurance was the medical insurance market for people over 65. Good risks do not buy this insurance because the average price is too high. He described other examples from the insurance literature and concludes that, “This principle of adverse

---

9 Optimal insurance coverage using a deductible was also analyzed by Pashigian et al. (1966) and Gould (1969).
10 Incomplete insurance may be obtained using a deductible or coinsurance (or both).
selection is potentially present in all lines of insurance” (page 493). He then discussed different market mechanisms put in place to reduce asymmetric information between insurance companies and insureds.

This article generated many contributions in the insurance literature. Rothschild and Stiglitz (1976) and Wilson (1977) presented conditions for obtaining a separating equilibrium, and Hoy (1982) and Crocker and Snow (1986) showed how risk classification, based on observable characteristics, can reduce adverse selection. Other significant contributions include Riley (1979) and Spence (1978).

2.5. Ehrlich and Becker (1972)

In the modern theory of risk management, insurance is seen as only one of the tools available to manage risk. The whole set of tools may be decomposed into subsets according to the different steps of the risk management process. Insurance belongs to the set of risk-transfer tools and represents a very powerful financial mechanism to transfer risk to the market. Another subset corresponds to risk prevention. Broadly, risk-prevention mechanisms may be classified under two headings: mechanisms intended to modify the probability of an event and mechanisms intended to mitigate the consequences of an event. Ehrlich and Becker (1972), in their article “Market Insurance, Self-Insurance, and Self-Protection,” published in the Journal of Political Economy, were the first to propose a rigorous economic analysis of risk prevention. They coined the terms self-protection and self-insurance to designate two kinds of activities and they studied their relationship to market insurance. For this reason, their paper may be seen as the first theoretical paper on risk management. Briefly, the paper provides three main results:

1) In the absence of market insurance, a risk-averse expected-utility maximizer will engage in self-protection and self-insurance activities, but the optimal investment in these activities depends on their cost. As usual, marginal benefit in terms of higher marginal expected
utility has to be weighted against the marginal disutility brought about by additional costs, so complete elimination of the risk is not optimal in general.

2) Self-insurance and market insurance are substitutes: an increase in the degree of protection provided by the insurer induces a rational individual to reduce their investment in activities (or behaviors) aimed at reducing the consequences of the insured event. Of course, this result is also important for the theory of moral hazard (see Section 4), but Ehrlich and Becker (1972) did not cover asymmetric information.

3) Self-protection and market insurance may be complements or substitutes, depending on the sensitivity of the insurance premium to the effects of self-protection. Thus, the insurer can give the insured an incentive to engage in self-protection activities (which reduce the likelihood of a loss) by introducing a link between the premium rate and the observation of such activities. This result is also important for the theory of moral hazard and, more generally, for agency theory (the theory of relationships between an informed agent and an uninformed principal).

2.6. Joskow (1973)

The article published by Paul Joskow in the *Bell Journal of Economics and Management Science*, under the title “Cartels, Competition and Regulation in the Property-Liability Insurance Industry,” was the first successful attempt to submit the insurance sector to an economic evaluation. The paper assesses competition by analyzing market concentration and barriers to entry. It measures returns to scale and discusses insurance distribution systems and rate regulation. By providing empirical results on these issues, it became a reference point for subsequent empirical research on the insurance sector. Briefly, Joskow (1973) found that the insurance industry was approximately competitive, that constant returns to scale could not be
excluded, and that the direct writer system was more efficient than the independent agency system.

The six seminal contributions presented above prepared the ground for numerous developments. That can be grouped under three main headings: the demand for insurance and protection, economic equilibrium under asymmetric information, and insurance market structure. They are addressed in sections 3 to 5. It is striking to realize that many of these developments are not limited to insurance economics per se. They occurred within the wider field of general economics, with insurance providing, in some cases, an illustration of general results and, in other cases, stimulation to search for general results.\(^{11}\)

### 3. Developments: Optimal insurance and protection

#### 3.1. The demand for insurance

Observations of economic life show that individuals generally do not insist on getting partial coverage when they take out an insurance policy. As the insurance premiums are generally loaded (at least to cover management costs), this is however the behavior that would be expected from the insured, according to Mossin’s (1968) results. Moreover, insurance does not seem to be empirically an inferior good, as would be implied by Mossin’s results, under DARA. If it was, insurance companies would be flourishing in the world’s poorer nations and would be among the declining industries in the richer ones. Recent empirical research on individual demand for insurance suggests that “higher income is (...) positively associated with insurance purchases of all kinds” (Cohen and Siegelman, 2010). This, again, contradicts Mossin’s analysis (given that absolute risk aversion is, indeed, empirically decreasing). One of

---

\(^{11}\) The survey of developments presented in the next three sections draws on the survey of insurance economics originally proposed by Dionne and Harrington (1992).
the seminal papers at the root of insurance economics thus led to two paradoxes. It is interesting to observe how theory was reconciled with factual observations. \(^{12}\)

The second paradox (insurance is an inferior good) did not stimulate much research effort. Some scholars tried to dig into the idea by exploring the conditions under which insurance would be not only an inferior good but also a Giffen good: see Hoy and Robson (1981), and Briys et al. (1989). But interest remained limited. There are probably two reasons for this. First, following Arrow (1970), it was quickly recognized among economists that insurance is a financial claim. Thus it does not really seem appropriate to apply to insurance concepts derived to categorize consumption goods. Second, it is likely that most scholars noticed that the condition under which Mossin’s result is obtained is not generally met in practice. Mossin assumes that the individual’s wealth increases, but that the risky component of wealth remains unchanged. In reality, changes in wealth generally imply changes in the portion of wealth exposed to a risk of loss, and this is sufficient to resolve the paradox (Chesney and Loubergé, 1986).

The first paradox (partial coverage is optimal) has stimulated much more research effort. It was first noticed that the result is not robust to changes in the pricing assumptions: for example, full insurance is optimal if the loading is a lump sum. \(^{13}\) Some researchers pointed out that the result was either strengthened or did not hold if the behavioral assumptions are modified: see Razin (1976) and Briys and Loubergé (1985), or the nonexpected utility developments (Machina, 1995). But the most interesting breakthrough came from broadening the scope of the analysis. This was done in the early eighties by deriving the logical conclusion from the observation that insurance is a financial claim. It had long been recognized

\(^{12}\) Other strange results were observed later on; for example, an increased loss probability has an ambiguous impact on insurance purchasing if the insured has DARA preferences and the insurer adjusts the premium to take the increased loss probability into account (Jang and Hadar, 1995).

\(^{13}\) It is obvious that the paradox can be resolved if one introduces differential risk perceptions. If the insured overestimates the probability (or the amount) of loss, full insurance may be optimal, even when the premium is loaded with a fixed proportional factor.
(Markowitz, 1959) that the demand for financial assets should take place in a portfolio context, taking into consideration correlations across random asset returns. The same kind of reasoning was applied to insurance by Mayers and Smith (1983), Doherty and Schlesinger (1983a, 1983b), Turnbull (1983), and Doherty (1984). In this portfolio approach, which was soon accepted as an important improvement, the demand for insurance coverage for one risk should not be analyzed in isolation from the other risks faced by the decision-maker: insurance demand is not separable, even when the risks are independent (Eeckhoudt and Kimball, 1992). When considering the insurance demand for one risk, one has to take into account the other risks, their stochastic dependence with the first risk, whether they are insurable or not and under what conditions, whether some insurance is compulsory or subsidized, whether a riskless asset is traded, and more: see, e.g., Schlesinger and Doherty (1985), von Schulenburg (1986), Kahane and Kroll (1985), Briys et al. (1988), and Gollier and Scarmure (1994). Thus, assuming that correlation is a sufficient measure of dependence, it may be optimal to partially insure a risk that is negatively correlated with another risk, even if the premium is actuarial. Conversely, it may be optimal to fully insure a risk despite unfair pricing if this risk is positively correlated with another uninsurable risk. In a portfolio context, incomplete markets for insurance provide a rationale for full insurance of the insurable risks. Mossin’s paradox can thus be resolved by changing the perspective rather than changing the analytical model (the expected utility model).

---

14 On a related theme, see also Doherty and Schlesinger (1990) for the case where the insurance contract itself is risky, due to a nonzero probability of insurer default. The paper shows that full insurance is not optimal under fair insurance pricing and that the usual comparative statics resulting from the single-risk model do not carry over to the model with default risk. Their work was extended by Cummins and Mahul (2003) to the case where the insurer and policyholder have divergent beliefs about the insurer default risk.

15 Hong et al. (2011) argued that correlation is not an adequate measure of stochastic dependence when expected utility is used. Turning to more general notions of positive and negative dependence, and focusing on coinsurance, they showed that the individual will purchase less than full (more than full) insurance if and only if the insurable risk is positively (negatively) expectation-dependent with random initial wealth.

16 These theoretical advances closely followed similar advances in the theory of risk premiums under multiple sources of risk: Kihlstrom et al. (1981), Ross (1981), Doherty et al. (1987). This literature on optimal insurance in presence of a background risk is closely linked to the literature on the demand for a risky asset, which was
Eeckhoudt and Kimball (1992) introduced the concept of prudence into the analysis of optimal insurance purchasing under background risk and pointed out that the demand for insurance for one risk was not independent of the background risk, even when the two risks are independent. Building on this premise, several papers derived the conditions under which optimal insurance demand under background risk has desirable comparative statics properties, such as an increase in optimal insurance coverage when the insured or uninsured risks increase, or whether a deductible policy remains optimal under background risk: see Meyer (1992), Dionne and Gollier (1992), Eeckhoudt et al. (1991, 1996), Gollier and Schlesinger (1995), Gollier (1995), Gollier and Pratt (1996), Tibiletti (1995), Guiso and Jappelli (1998), Meyer and Meyer (1998a), Mahul (2000), and Gollier and Schlee (2006).

The insurance model with background risk has been extended to the case where the uninsurable risk is nonpecuniary. This is the case, for example, if the background risk represents a state of health. The problem may be analyzed using state-dependent utility functions or introducing a second argument in the decision-maker’s utility function, besides wealth. Using the second approach, Rey (2003) demonstrated that the impact of the nonfinancial risk on insurance demand depends not only on the relationship between the two risks but also on the impact of the background risk on the marginal utility of wealth. For instance, if the marginal utility of wealth increases under occurrence of a nonfinancial loss (some degree of disability, for example) and if the two risks are positively correlated, then insurance demand will increase. Full insurance becomes possible, even with a loaded premium. But full insurance with a loaded premium may also be obtained if the two risks are negatively correlated and the

pioneered by Arrow (1963) in a single-risk setting and developed later to consider the impact of background risks: see, e.g., Tsetlin and Winkler (2005) and Li (2011).

17 This case corresponds to a negative cross-derivative of the two-attribute utility function ($u_{12} \leq 0$). Eeckhoudt et al. (2007) showed that this is equivalent to “correlation aversion,” i.e., the aversion to losses simultaneously affecting the two attributes of utility (health and wealth, for example).
marginal utility of wealth is lower under occurrence of a nonfinancial loss. The range of possibilities for contradicting Mossin’s first proposition (in 2.3 above) expands.

3.2. Insurance, consumption, and saving

Research integrating joint optimal decisions on consumption, saving, and insurance each represents a different research program, which was addressed by Moffet (1977) and Dionne and Eeckhoudt (1984). The latter paper showed that investing in a riskless asset is a substitute for purchasing insurance. This work was generalized by Briys (1988), using a continuous-time model. Gollier (2003) considered the impact of time diversification on optimal insurance demand in a dynamic framework, under the assumption of no serial correlation in risks. He showed that liquidity-constrained individuals will insure widely. Wealthy individuals will take advantage of time diversification to accumulate buffer stock wealth and avoid the costs due to loaded insurance prices.

A related avenue of research concerns the joint determination of insurable asset purchases and optimal insurance coverage: see Meyer and Ormiston (1995), Eeckhoudt et al. (1997), Meyer and Meyer (2004), and Loubergé and Watt (2008) for work along this line. In the first of these papers, the individual is endowed with riskless wealth and with a risky insurable asset, but insurance can only be paid for by selling a portion of the risky asset. Hence, the model differs from that of Mossin (1968), where insurance is paid out of riskless wealth. However, because an increase in wealth impacts only riskless wealth, insurance remains an inferior asset under DARA. Eeckhoudt et al. (1997) generalized the previous work by considering an individual who allocates some nonrandom wealth to the purchase of a safe asset, a risky asset, and an insurance contract to cover the risky asset. The optimal demands for the risky asset and

---

18 Following Eeckhoudt et al. (2007), the individual is then “correlation loving.” For that individual, in this case, purchasing more insurance against a loss in wealth helps to mitigate the adverse impact of a negative correlation between the two risks.

19 The following developments on this topic borrow from the literature review in Loubergé and Watt (2008).
for insurance coverage are determined jointly and paid for out of risk-free wealth. As the insurance indemnity is not linear,\(^{20}\) it may be optimal to keep some wealth in the risk-free asset. As expected, it turns out that insurance and holding a riskless asset are substitutes. However, the generality of the model did not allow the authors to derive clear comparative statics results. In particular, “increases in initial wealth can lead to increases or decreases in the insurance level and to increases or decreases in the holding of the risky asset, (...) even when the decision-maker is decreasingly risk averse” (Eeckhoudt et al., 1997, p. 26). Thus Mossin’s second paradox is not confirmed: insurance is not necessarily inferior under DARA. Further, Meyer and Meyer (2004) considered the case where the individual is endowed with a composite portfolio of risky and riskless assets in fixed proportion. The risky asset may be insured without changing the proportions of the two assets held. Under these peculiar circumstances, insurance turns out to be normal whenever relative risk aversion is nondecreasing. In addition, the authors proved that insurance is ordinary (not Giffen) if relative risk aversion is less than or equal to one, a condition already derived in previous work based on the standard model (Hoy and Robson, 1981). The same restrictive condition is also pivotal—as in Meyer and Ormiston (1995) and in Eeckhoudt et al. (1997)—to determining whether or not insurance demand will increase in reaction to an increase in the size of the loss. Loubergé and Watt (2008) addressed the same issues by focusing on the case where the riskless asset is dominated and all available wealth is invested in a risky and insurable asset (an investment opportunity), partly to purchase the asset, partly to finance the purchasing of insurance. In their setting, coinsurance is allowed, and partial insurance is optimal when the premium is loaded and may be optimal when insurance is fair. The fraction of the investment subject to a loss is a very important parameter in the model, along with risk aversion. If the fraction is low enough, no insurance is optimal. With a larger fraction, and positive insurance, insurance is normal if relative risk aversion is

\(^{20}\) Coinsurance must be excluded to avoid corner solutions of either no holding of the riskless asset or zero demand for insurance. Meyer and Meyer (1998) address the specific case of deductible insurance.
nondecreasing. But even with decreasing relative risk aversion, insurance is still normal if the proportion of the investment subject to a loss is higher than the rate at which relative risk aversion decreases. Insurance demand increases unambiguously if the percentage loss increases. But, when the loss probability increases and the insurer simultaneously adjusts the premium rate to take this change into account, the results are less clear-cut: it turns out that the demand for insurance increases if relative risk aversion is constant and less than or equal to 1; but it may also increase for values of constant relative risk aversion larger than 1, depending on the relationship between this value and the value of the percentage loss to which the asset is exposed.Obviously, changes in risk aversion complicate the analysis when purchasing an insurable risky asset and purchasing insurance are considered simultaneously instead of separately.

3.3. Self-protection and self-insurance

During the eighties, research on risk prevention (self-protection and self-insurance activities, in the sense of Ehrlich-Becker, 1972) developed more slowly than research on insurance demand, but it received increased attention thereafter. The earlier contributions came from Boyer and Dionne (1989a), who noted that self-insurance leads to stronger changes in risk than self-protection (see also Chang and Ehrlich, 1985), and from Dionne and Eeckhoudt (1985), who showed that increased risk aversion leads to more self-insurance, but obtained the surprising result that an increase in risk aversion does not necessarily result in higher self-protection, everything else being constant. Briys and Schlesinger (1990) later proved that this result is quite robust to a change in the model setting, e.g., introducing state-dependent utility

\[ \text{21 In this case, with increasing wealth, the rate of increase of the possible loss amount is higher than the rate of decrease of relative risk aversion.} \]

\[ \text{22 Using Mossin’s (1968) approach, Jang and Hadar (1995) obtained that the effect of an increase in the probability of loss is, in this case, indeterminate if the utility function displays DARA, and that the demand for insurance decreases with a CARA or IARA utility.} \]

\[ \text{23 See Schlesinger (2013) for a survey on insurance demand.} \]
functions or a random initial wealth. As written by Briys and Schlesinger (1990), “an increase in the level of self-protection causes both a mean-preserving spread and a mean-preserving contraction in the wealth distribution, with the spread occurring at lower wealth levels and the contraction at higher wealth levels” (p. 465). For this reason, a more risk-averse individual does not necessarily invest more in self-protection. Eeckhoudt and Gollier (2005) showed that introducing prudence into the picture does not lead to a more intuitive result. More prudence (in the sense of Kimball, 1990) does not lead to more self-protection. In particular, if the optimal self-protection expenditure of a risk-neutral agent is such that their probability of loss is $p_n \geq \frac{1}{2}$, a risk-averse and prudent agent spends less on self-protection than does the risk-neutral agent (see also Dachraoui et al., 2004, and Dionne and Li, 2011, for comparable results). The reason is that prevention has a current monetary cost (it is defined as an expenditure) and a more prudent individual wants to increase savings to hedge against future contingencies.

Courbage et al. (2017) replaced the analysis of prevention in the broader multiple-risks context. Steps in that direction had already been taken by Briys and Schlesinger (1990) and by Briys et al. (1991) with their analysis of risky risk management. Later, Dachraoui et al. (2004) noted that their analysis of self-protection for a mixed risk-averse agent à la Caballé and Pomansky (1996) also applies when the agent faces a background risk. An important additional step on this issue was also taken by Lee (2012) in his analysis of self-protection under background risk. He considered two kinds of self-protection: self-protection effort and monetary investment in self-protection devices. He obtained that an individual facing a background risk will exert more self-protection effort than the same individual without

---

24 See also Courbage and Rey (2006) for an extension of this result to the case of two-argument utility functions, wealth and health.

25 The recent models are becoming very complex without generating less ambiguous results (Peter, 2021a). A probability of loss equal to (greater than, lower than) 0.5 remains a crucial threshold for comparative statics analysis because it affects the direction of the variation of the variance and other statistical moments when we change the probability of loss in a two-state model. Moreover, it is not dependent on decision-makers’ risk behavior.
background risk. Concerning monetary investment in self-protection, Lee was also able to show that the presence of a background risk will increase self-protection if the self-protection expenditure is paid out of income, not wealth, and if wealth and consumption are complements.26

Following the lead of Menegatti (2009), who introduced a two-period model to consider optimal prevention, Courbage and Rey (2012) investigated the impact of background risks on optimal self-protection expenditures in a two-period model. They showed that a prudent individual does not necessarily exert more effort in the presence of background risk. The results differ depending on whether the background risk is introduced in the first or second period, and depending on whether the background risk arises in the loss or no-loss state of nature. Eeckhoudt et al. (2012) also used a two-period model to determine the impact of a background risk on protection expenditure. In their model, the background risk appears in the second period and it increases prevention under prudence. Moreover, a deterioration of the background risk generates still more prevention under mixed risk aversion.27

The model proposed by Courbage et al. (2017) is more general than the previous works because it does not assume a “background risk.”28 There are, at least, two risks, and the risks may be positively or negatively correlated, or independent. The authors “wonder how the characteristics of the multiple risks and the efficiency of alternative prevention instruments influence the decision to invest in a portfolio of prevention measures” (p. 900). They showed that introducing a second, possibly correlated, risk increases prevention for the first risk, independently of the sign of correlation. The prevention expenditure for the first risk also

26 Note that all this literature on self-insurance and self-protection has been driving away from the study of the links between insurance demand and prevention. In addition, except for the paper by Lee (2012), it has focused on the case where prevention implies a monetary cost (prevention expenditures), instead of the case where prevention implies an effort producing a direct loss in utility, because the analysis of the latter case is more straightforward even if it is often applied in principal-agent models.
27 See also Wang and Li (2015).
28 But, like previous work, it deals with self-protection, leaving self-insurance aside.
increases if the two risks become more interdependent or if the severity of the second risk increases (higher expected loss). When the two risks are endogenous, a substitution effect arises in the optimal prevention for each risk, with important public policy implications: exogenously raising the required prevention level for one risk leads to a drop in the optimal prevention for the other risk.

In addition, Hofmann and Peter (2016) found that the presence of savings also matters for the effect of utility curvature on self-insurance and self-protection. As shown by Peter (2017), in general, the presence of savings matters for the comparative statics results of two-period self-protection.

Peter (2024a, 2024b) presented two detailed surveys that summarize the different technical difficulties in obtaining intuitive results in existing models on self-protection research. He emphasized the role of probability thresholds in obtaining more precise comparative statics results and reviewed extensions to nonexpected utility and recent contributions in two-period models.²⁹

Already Ehrlich and Becker (1972) had provided the main intuition to understand the self-protection puzzle. They noticed that, unlike insurance, self-protection does not redistribute income among world states. Rather, the amount spent to decrease the probability of loss reduces final wealth in all states, leaving the size of the loss unchanged. Peter (2024a) presented an example demonstrating the importance of this feature of the self-protection puzzle. The main result is that the comparative statics of self-protection do not correspond to the single-crossing property proposed by Jewitt (1989) for comparative statics analysis under uncertainty.

Jullien et al. (1999) presented a method to understand this (see also Lee, 1998). Considering two individuals with different levels of risk aversion, they showed that an increase

²⁹ See also the Geneva Risk Economics Lecture presented by Bleichrodt (2022) for another review with an emphasis on nonexpected utility. He also discusses how empirical research may help to understand the prevention puzzle.
in risk aversion raises the demand for self-protection if and only if the loss probability of the less risk-averse decision-maker (DM) is small enough. The use of probability threshold results like the one discovered by Jullien et al. (1999) became common in the economic analysis of self-protection. However, the optimal loss probability is endogenous because it depends on the risk preferences of the reference DM’s utility function. The probability threshold itself is also an endogenous quantity because it depends on the DM’s utility function and the risk preferences of the reference DM. The main question is how to determine the magnitude of the threshold in a given application (Dachraoui et al., 2004). How to measure this threshold empirically is not obvious!

Turning to a different question, Dionne and Eeckhoudt (1988) analyzed the effect of increasing risk on self-protection. In particular, they studied a marginal increase in the loss amount that is compensated by an increase in initial wealth such that the expected final wealth stays constant (mean-preserving spread). Counterintuitively, such a change again has no definitive effect on self-protection and may well lead to a decrease in its optimal level. Peter (2024a) extended the problem raised by Dionne and Eeckhoudt (1988) and showed how the use of threshold analysis generates definitive predictions.

Sweeney and Beard (1992) studied the wealth effects on self-protection. Unlike insurance, DARA is neither necessary nor sufficient for an increase in wealth to reduce the optimal demand for self-protection. Again, the effect depends on the loss probability relative to a threshold value. Peter (2022) revisited the question of wealth effects and presented two main findings. First, the probability threshold can be calibrated for various classes of utility functions. These parametric results suggest that self-protection is an inferior good. Self-protection can even be Giffen, which is, however, not empirically plausible, according to Peter (2021b). Second, when considering that the value of assets subject to prevention is increasing in income, self-protection is more plausibly a normal good.
One of the biggest shortcomings of this literature is the lack of empirical work on self-protection. Courbage et al. (2013) made this point ten years ago already, and little progress has been observed since then. For example, prevention in health should be studied more. Courbage and Rey (2006) and Peter (2021a) considered self-protection against a health risk, which is often referred to as primary prevention in health economics. Another field of application should be climate risk. Insurance companies are starting to leave this market, which is considered too risky. How private and social self-protection activities can be used for population resilience will become a very important issue in the coming years. However, it is not clear that the theoretical results in the actual self-protection literature would be very useful. More theoretical research is still needed, with careful attention given to applicability.

Another problem with this literature is the use of different endogenous thresholds, which depend on the nature of the utility function, to obtain intuitive comparative statics results. This is quite problematic for tests in empirical studies. It is thus very difficult for a public authority to implement preventive projects for a population with different nonobservable risk attitudes and substitution effects between prevention efforts for different risks.

3.4. The demand for liability insurance

Liability risk raises particular issues that have been addressed in a specific branch of the insurance economics literature, at the interface between law and economics. An economic agent (the injurer) may be made liable for the monetary and nonmonetary losses they impose on another agent (the victim). The losses are random but are generally influenced by the injurer’s decisions, regarding his/her level of potentially harmful activity, and level of prevention. The injurer can contract liability insurance to cover the risk of being sued by the victim(s). The availability of insurance is not without influence on the injurer’s decisions regarding the levels of activity and care, as has been known since Ehrlich and Becker (1972), but the specificity of
liability insurance arises from the possibility that the losses imposed on the victim(s) exceed the injurer’s wealth. In this case, the injurer is judgment proof and cannot be forced to bear the full monetary consequences of the losses resulting from his/her activity. This influences the injurer’s optimal level of care and insurance demand.

The impact of the judgment-proof problem on the demand for liability insurance was first analyzed by Sinn (1982). He remarked that when injurers are socially guaranteed a minimum subsistence level of wealth, a kink appears in their utility function. The kink breaks the overall concavity of the function, with the result that a risk-averse injurer may rationally choose not to purchase any liability insurance even if the insurance premium is actuarially fair. Whether or not insurance will be purchased depends on the injurer’s initial wealth, on the socially guaranteed minimum subsistence level, and on the size of the loss, among other usual influences such as risk aversion and insurance price. If insurance is purchased, it will be full insurance if the premium is actuarially fair (no proportional loading applies), but partial insurance if the premium entails proportional loading—as would be expected from Mossin’s (1968) results. However, a noteworthy implication of the analysis is that insurance demand for liability insurance is an increasing function of the injurer’s initial wealth, even if the injurer’s preferences are DARA and the possible loss is fixed—a result opposite from the one obtained by Mossin in the property insurance case.

Huberman et al. (1983) emphasized that the reluctance of injurers to purchase liability insurance derives from the fact that the insurance premium takes into account a range of (high) losses to which injurers are not exposed if such losses exceed the value of their assets (or the difference between their assets and the socially guaranteed minimum subsistence level). Using the example of a three-state model and assuming actuarially fair insurance pricing, they showed that when a possible large loss exceeds the injurer’s assets, it is preferable for the injurer to set
a limit on the insurance indemnity and remain partially covered, even for losses that do not imply bankruptcy. The risk-spreading function of insurance is hampered.

Liability insurance and the judgment-proof problem were comprehensively analyzed by Shavell (1986), in a model where the potential injurer decides simultaneously on whether to purchase insurance and on the optimal level of care, under two possible legal frameworks: strict liability and the negligence rule. Under strict liability, the results depend on whether or not the insurer can observe the insured’s level of care. With perfect information, the insurer can adjust the premium to the level of care. In this case, if the injurer’s initial wealth exceeds some threshold, full insurance and an efficient level of care are both optimal. As care is observable and impacts the insurance premium, the insured gains from adopting the efficient level of care, whereas in the absence of insurance, a lower level of care would have been optimally chosen, due to the judgment-proof problem. If the injurer’s initial wealth is below the threshold, no insurance is purchased and the level of care is zero or reduced below the efficient level.

These results make a strong case for imposing compulsory liability insurance but they are conditioned on the assumption of perfect observability. If care exercised by the injurer is not observable by the insurer (moral hazard), either no insurance is chosen (in a wider range of injurer’s wealth) or insurance is partial, with a sufficient level of care. In this situation, the case for compulsory liability insurance is not made. A tension arises between risk-spreading and appropriate incentives to avoid losses.

When the negligence rule holds, care is assumed to be observable. In this case, it is optimal for the injurer not to purchase any insurance and to exercise the optimal no-insurance level of care when the judgment-proof problem arises: either no care at all if initial wealth is

---

30 Under the negligence rule, an injurer cannot be made liable for the losses imposed on a victim if the injurer has applied the appropriate level of care. In theoretical work, the appropriate level of care is the socially efficient level of care optimally chosen by a risk-averse potential injurer if the judgment-proof problem does not arise. This level balances the marginal benefits and marginal cost of care.
below a first threshold; an increasing level of care if initial wealth is beyond this threshold but below a second threshold; and the efficient level of care if initial wealth is beyond the second threshold. The reason is that it is useless to purchase insurance if the injurer applies the appropriate level of care. Not applying this level and purchasing coverage for the risk of being liable would be more costly.\footnote{In addition, the injurer would run the risk of being denied indemnification by the insurer if it turned out \textit{ex post} that the level of care had been inappropriate.} Of course, the result under the negligence rule hinges on the belief that the injurer will not be judged liable for the loss imposed on the victim(s) if the efficient level of care was chosen. Judicial uncertainty is ruled out (see Shavell, 2000, for more on this issue). More generally, the results presented above rely on the assumption that the legal tort liability / liability insurance system is efficient, which has been hotly debated, particularly in the wake of the US liability insurance crisis of the mid-1980s. (See Danzon and Harrington, 1992, for an early survey of liability insurance issues; and Ambrose et al., 2013, Baker and Siegelman, 2013, Faure, 2014, and Shavell, 2018, for more recent surveys). Parra and Winter (2024) discussed the interaction of protection against risk through limited liability and protection through insurance. Incentives from tort law, in presence of moral hazard, was also addressed. The literature on liability insurance has not been especially dynamic in recent years. This may be explained by a few changes in laws and regulations.

3.5. Other contributions

Other work in the theory of optimal insurance has concerned the following:

1) The specific issues raised by the corporate demand for insurance: these issues will be considered in Section 6 below.

2) The focus on anomalies observed in actual insurance purchasing behavior (Kunreuther and Pauly, 1985).
3) The extension of the expected utility model to take into account state-dependent utility functions. One can thus introduce into the analysis important observations from reality. For example, the observation that the indemnity paid by the insurer cannot provide complete compensation for a nonmonetary loss, such as the loss of a child, or the observation that the marginal utility of wealth is different under good health than under disability: see Arrow (1974), Cook and Graham (1977), and Schlesinger (1984) for important papers along this line. This line of research is related to, and generalized by, the literature on insurance purchasing under multiattribute utility functions (Rey, 2003).

4) The replacement of the expected utility model with recent generalizations, grouped under the heading “nonexpected utility analysis.” This research program has produced several interesting results. Using the distinction between risk aversion of order 1 (first-order risk aversion) and risk aversion of order 2 (second-order risk aversion), Segal and Spivak (1990) showed that Mossin’s (1968) result on the optimality of partial coverage under a loaded insurance premium does not necessarily hold if risk aversion is of order 1 (see also Schlesinger, 1997). Now, risk aversion of order 1 may occur under the expected utility model (if the utility is not differentiable at the endowment point) or under some generalizations of this model, such as Yaari’s (1987) dual theory or Quiggin’s (1982) rank-dependent expected-utility theory, and in the presence of a dependent background risk (Dionne and Li, 2014). In particular, using Yaari’s model, Doherty and Eeckhoudt (1995) showed that only full insurance or no insurance (corner solutions) are optimal with proportional insurance when the premium is loaded. Karni (1992) showed that Arrow’s (1963) result on the optimality of a deductible policy is robust to a change in behavioral assumptions if the modified model satisfies some differentiability conditions, which are met by Yaari’s (1987) and Quiggin’s (1982) models. Konrad and

---

32 The orders of risk aversion, as defined by Segal and Spivak (1990), rest on the behavior of the risk premium in the limit, as the risk tends toward zero.

33 This result is reminiscent of the same result obtained under Hurwicz’s model of choice under risk: see Briys and Loubergé (1985).
Skaperdas (1993) applied Ehrlich and Becker’s (1972) analysis of self-insurance and self-protection to the rank-dependent expected utility model. Schlee (1995) confronted the comparative statics of deductible insurance in the two classes of model. To date, the most comprehensive attempt to submit classical results in insurance economics to a robustness test by shifting from expected utility to nonexpected utility can be found in Machina (1995). Using his generalized expected utility analysis (Machina, 1982) he concluded that most of the results are quite robust to dropping the expected utility hypothesis. However, the generality of his conclusion was challenged by Karni (1995), since Segal and Spivak (1990) showed that Machina’s generalized expected utility theory is characterized by risk aversion of order 2.

The demand for insurance under background risk in a nonexpected utility setting was analyzed by Doherty and Eeckhoudt (1995) using Yaari’s (1987) dual choice theory. They showed that partial insurance may be obtained under proportional coverage and a loaded insurance premium if an independent background risk is present (full insurance remains optimal if the premium is fair). Dropping the independence assumption, they noted that the likelihood of getting a corner solution increases. But, qualitatively, the effects of introducing positively or negatively correlated background risks are the same as under expected utility. More generally, Schlesinger (1997) showed that introducing an independent background risk in a decision model with risk aversion does not change the predictions obtained under a single source of risk: full insurance is optimal under a fair premium; partial or full insurance may be optimal under a loaded premium; and a deductible policy remains optimal.
4. Developments: Markets under asymmetric information

The Arrow (1953) model shows that a market economy leads to a general and efficient economic equilibrium—even under risk and uncertainty—if the financial market is complete, i.e., provided the traded securities and insurance contracts make it possible to optimally cover any future contingency. This is an important result since it extends to the case of risk and uncertainty the classical result on the viability and efficiency of a competitive market economy.

However, as Arrow himself noticed in his 1963 article, full insurance coverage is not always available in insurance markets, for various reasons. Among these reasons, asymmetric information has received much attention in the economic literature and has been generally discussed under two main headings: moral hazard and adverse selection. Moral hazard exists when (1) the contract outcome is partly under the influence of the insured; and (2) the insurer is unable to observe, without costs, to what extent the realized losses are attributable to the insured’s behavior. Adverse selection occurs when (1) the prospective insureds are heterogeneous, and (2) the risk class to which they belong cannot be determined \textit{a priori} by the insurer (at least not without costs), so that every insured is charged the average premium rate. Clearly, asymmetric information is a source of incompleteness in insurance markets, e.g., a student cannot be insured against the risk of failing an exam; a healthy old person may not find medical insurance coverage at an acceptable premium, etc. For this reason, a competitive

\footnote{An economic equilibrium is first-best efficient if it is Pareto-optimal: it is impossible to organize a reallocation of resources that would increase the satisfaction of one individual without hurting at least one other individual. The first theorem of welfare economics states that any competitive equilibrium is Pareto-optimal, and the second theorem states that a particular Pareto optimum may be reached by combining lump sum transfers among agents within a competitive economic system. In an efficient equilibrium, market prices reflect social opportunity costs.}

\footnote{In an interesting article on the history of the term “moral hazard,” Rowell and Connelly (2012) noted that the concepts of moral hazard and adverse selection have often been confused in the traditional insurance literature. They also remarked that this literature tends to attribute a pejorative meaning to “moral hazard,” often associating it with fraud, in contrast to the economic literature, which focuses on incentives and maintains that “moral hazard has in fact little to do with morality” (Pauly, 1968).}
market economy may not be first-best efficient, and this may justify second-best resource allocation mechanisms and even government intervention.

4.1. Moral hazard

Economists make a distinction between two kinds of moral hazard, depending on the timing of the insured’s action. If the unobservable action occurs before the realization of the insured event, one has *ex ante* moral hazard, while *ex post* moral hazard exists when the insured’s action is taken after the insured event.  

*Ex ante* moral hazard has been studied by Pauly (1974), Marshall (1976), Holmstrom (1979), and Shavell (1979), among others. They showed that insurance reduces the incentive to take care when the insurer is unable to monitor the insured’s actions. More generally, partial provision of insurance is optimal under moral hazard. More specifically it was demonstrated that actuarial pricing is not optimal when the insured’s behavior affects the probability of a loss. The equilibrium premium rate is an increasing function of the amount of coverage purchased (nonlinear pricing, as in Pauly, 1974). In addition, under moral hazard in loss reduction, the optimal contract is conceived such as to make the degree of coverage a nonincreasing function of the amount of losses, with large losses signalling careless behavior by the insured. Small losses are fully covered, but losses exceeding a limit are partially covered (Parra and Winter, 2024). In a security design analysis, Holmstrom (1979) presented conditions under which a deductible or a coinsurance rate above a deductible is optimal (see also Shavell, 1979). Shavell

---

36 *Ex post* moral hazard is particularly important in medical insurance, where claimed expenses are dependent on decisions made by the patient and the physician once illness has occurred. It is also important in workers’ compensation and in auto bodily injury insurance.

37 Dionne (1982) pointed out that moral hazard is also present when the insured event results in nonmonetary losses, for example the loss of an irreplaceable commodity.
(1982, 1986) extended the study of moral hazard to the case of liability insurance. He showed that making liability insurance compulsory results in less than optimal care.\(^{38}\)

The existence of long-term (multiperiod) contracts does not necessarily mitigate the effect of moral hazard. Under the infinite-period case, Rubinstein and Yaari (1983) proved that the insurer can eliminate the moral hazard problem by choosing an appropriate experience rating scheme that provides full incentives to take care. But the result does not, in general, carry over to the finite-period case (Parra and Winter, 2024). In addition, the possibility for the insured to switch to another insurer makes a penalty scheme difficult to enforce in competitive insurance markets, where insurers do not share information about prospective insureds.\(^{39}\)

*Ex post* moral hazard was first pointed out by Spence and Zeckhauser (1971) and later studied by Townsend (1979). In this case, the nature of the accident is not observable by the insurer, who has to rely on the insured’s report or engage in costly verification.\(^{40}\) Mookherjee and Png (1989) showed that random auditing is the appropriate response by the insurer in this situation. Their work was extended by Fagart and Picard (1999), who investigated the characteristics of optimal insurance under random auditing. Using a deterministic auditing policy, Bond and Crocker (1997) obtained that the optimal insurance contract includes generous payments of easily monitored losses and undercompensation for claims exhibiting higher verification costs.

The consequences of moral hazard for the efficiency of a market economy were studied by Helpman and Laffont (1975), Stiglitz (1983), Arnott and Stiglitz (1990), and Arnott (1992),

\(^{38}\) Note that moral hazard is also present in the insurer–reinsurer relationship (see Jean-Baptiste and Santomero, 2000). The success of index products in insurance securitization is partly due to the fact that they remove the moral hazard from the relationship between insurers and providers of reinsurance coverage (Doherty and Richter, 2002).

\(^{39}\) The situation is of course different in regulated markets where information on retrospective rating is shared by the market.

\(^{40}\) At some point, this moral hazard problem becomes a fraud problem. See Picard (1996), Crocker and Morgan (1998), the special issue on fraud in *The Journal of Risk and Insurance*, September 2002 (Derrig, 2002), and more recently Dionne et al. (2009). See Picard (2024) for a survey of the literature.
among others. They showed that a competitive equilibrium may not exist under moral hazard, and that the failure to get complete insurance coverage results at best in a subefficient equilibrium. This is due to the fact that moral hazard involves a trade-off between the goal of efficient risk bearing, which is met by allocating the risk to a less risk-averse insurer, and the goal of efficient incentives, which requires leaving the consequences of decisions about care to the insured (Parra and Winter, 2024). However, government intervention does not necessarily improve welfare in this case. This depends on government information, compared with the information at the disposal of private insurers. Arguments may be put forward in favor of a taxation and subsidization policy providing incentives to avoid and reduce losses, but public provision of insurance does not solve the moral hazard problem (Arnott and Stiglitz, 1990).

Moral hazard has become a popular theme in economics, not only because its presence in insurance markets results in less than optimal functioning of any economic system, but also because it is a widespread phenomenon. As Parra and Winter (2024) noted, moral hazard can be broadly defined as a conflict of interests between an individual (behaving rationally) in an organization and the collective interest of the organization. Insurance markets provide the best illustration for the effect of moral hazard, but this effect is also observed in labor relationships, in finance contracts, and quite generally in all circumstances where the final wealth of a principal is both uncertain and partially dependent upon the behavior of an agent whose actions are imperfectly observable. For example, in a corporation, the wealth of the firm’s owners (stockholders) is partly dependent upon the actions of the manager; in judicial procedures, the final outcome is partly dependent upon the efforts of the lawyers; in a team, the success of the team is partly dependent on the individual effort of the members; and, in a bank loan, the default state is dependent on the borrower’s behavior. All these situations have been studied in the economic and financial literature under the headings of principal–agent relationships or agency theory, with close connections to the literature on moral hazard in insurance. In both cases, the objectives are to define the optimal incentive contract to mitigate the effect of asymmetric
information and to study the consequences of different arrangements on deviations from efficiency: see Ross (1973), Radner (1981), Lambert (1983), and Grossman and Hart (1983) for canonical references. See also Allen (1985), Fudenberg and Tirole (1990), and Chiappori et al. (1994) for research introducing credit markets and savings into the analysis. Similarly, the consequences for general economic equilibrium of market incompleteness brought about by moral hazard, among other causes, has become a central theme of research in economics: see, e.g., Polemarchakis (1990). Moral hazard developments in insurance economics were closely related to developments in general economic theory.

Turning to empirical work, and focusing on moral hazard in insurance contracts, evidence for moral hazard impacting on insured losses has been documented in several studies, taking advantage of natural experiments provided by changes in legislation, starting with Dionne and St-Michel (1991) for workers’ compensation, and continuing with Boyer and Dionne (1989b) and Cohen and Dehejia (2004) for automobile insurance, and with Chiappori et al. (1998) and Klick and Stratmann (2007) for health insurance. More recently, Dionne and Liu (2021) applied a difference-in-differences methodology to separate moral hazard from adverse selection in an experiment in China where the bonus-malus was modified in a city and not modified in another similar city. They did not reject the presence of moral hazard while controlling for adverse selection. By contrast, Abbring et al. (2003) did not find any evidence of moral hazard in multiperiod data provided by the French system of bonus-malus in automobile insurance. A malus, which results from prior accident history and increases the cost of insurance for the policyholder, did not lead to a significant reduction in insured losses in this study.

Dionne et al (2013) extended this literature by proposing a dynamic methodology to separate moral hazard from adverse selection (and asymmetric learning). Since there are two potential information problems in the data, a simple correlation between risk and insurance coverage is not sufficient to isolate the causality between risk and insurance coverage. They
added a contract choice equation to the standard claim equation in their dynamic econometric model. They did not reject moral hazard and asymmetric learning for certain groups of insured.

In addition, even if a drop in insurance claims is observed following the introduction of an experience rating, evidence from Canada suggests that much of the decline is due to an increased incentive not to report small claims to avoid premium increases (see also Robinson and Zheng, 2010). In this case, policy changes introduced to address the *ex ante* moral hazard issue stimulate the development of a kind of *ex post* moral hazard. In the case of fraud—an extreme version of moral hazard—Hoyt et al. (2006) reported that antifraud laws introduced in the United States in the period between 1988 and 1999 had mixed effects on automobile insurance fraud. Some laws had no statistically significant effects, while others actually increased fraud. More recently, Dionne et al. (2009) found significant *ex post* moral hazard in the portfolio of a European insurer.

To sum up, although there are strong theoretical reasons to believe that moral hazard is a major issue in insurance and that experience rating is a powerful tool for dealing with it, the empirical evidence so far is not compelling. This is probably due to the difficulty for econometricians to set up tests that isolate moral hazard from other influences on accidents, given the information limits on actual incentives and behavior among insureds. As pointed out by Parra and Winter (2024), the real incentives of repeated insurance contracts through the threat of higher future premiums still deserve more thinking.

4.2. Adverse selection

A central development in the study of adverse selection was the paper by Rothschild and Stiglitz (1976). This paper assumed two classes in the insured population: good risks and bad risks. The two classes differ only with respect to their accident probability. The authors showed that a competitive insurance market does not reach a pooling equilibrium under adverse
selection. Moreover, under the assumptions of the model, including the assumption of a pure Cournot-Nash strategy by insurers, a separating equilibrium is obtained if the proportion of good risks in the economy is not too large. The equilibrium situation involves the supply of discriminating contracts providing full insurance at a high price to the bad risks and partial coverage at a low price to the good risks.\textsuperscript{41} Compared to the symmetric information case, the bad risks get the same expected utility but the good risks suffer a welfare loss. The policy implication of the model is that, in some circumstances, insurance markets may fail, and monopolistic insurance (under government supervision) or compulsory insurance may be justified as a second best.\textsuperscript{42}

Extensions of the basic Rothschild-Stiglitz model are due to Wilson (1977), Spence (1978), and Riley (1979), who dropped the assumption of a Cournot-Nash strategy by insurers. Then, an equilibrium always exists, either as a separating equilibrium (Riley, Wilson) or as a pooling equilibrium (Wilson). Moreover, Spence (1978) showed that this equilibrium is efficient if the discriminating insurance contracts are combined with cross-subsidization among risk classes, with the low risks subsidizing the high risks.\textsuperscript{43} Extensions have concerned the case where individuals face a random loss distribution (Doherty and Jung, 1993; Doherty and Schlesinger, 1995; Landsberger and Meilijson, 1996; Young and Browne, 1997), the case where they differ with respect to both accident probability and degree of risk aversion (Smart, 2000; Villeneuve, 2003), the case where some of them are overconfident (Sandroni and Squintani, 2007), and the case where they are exposed to multiple risks or background risk (Fluet and Pannequin, 1997; Crocker and Snow, 2008). Allard et al. (1997) have also shown

\textsuperscript{41} Insurance contracts are defined in terms of price and quantity, instead of price for any quantity. Insureds reveal their class by their choice from the menu of contracts, such as the choice of a deductible.

\textsuperscript{42} Stiglitz (1977) studied the monopolistic insurance case. Under asymmetric information, the monopolist insurer maximizes profit by supplying a menu of discriminating contracts. At the equilibrium situation, the high risks get some consumer surplus, but the low risks are restricted to partial insurance and do not get any surplus.

\textsuperscript{43} See Crocker and Snow (1985) for a review of these models, and Dionne et al. (2024) for a survey of adverse selection theory.
that the Rothschild-Stiglitz results are not robust to the introduction of transaction costs: for arbitrary small fixed setup costs, pooling equilibria may exist in a competitive insurance market, and high-risk individuals (rather than low-risk individuals) are rationed. In addition, it is important to note that a separating equilibrium may be invalidated if insureds have the opportunity to purchase coverage for the same risk from different insurers. For this reason, Hellwig (1987, 1988) presented models to take into account the sharing of information by insurers about the policyholders. Mimra and Wambach (2019) extended these models to introduce dynamic strategic interaction in competitive markets.

Adverse selection was empirically tested by Dahlby (1983, 1992) for the Canadian automobile insurance market, and by Puelz and Snow (1994), who used individual data provided by an automobile insurer in the US state of Georgia. Both studies reported strong evidence for adverse selection and provided empirical support for the separating equilibrium outcome; in addition, the former study found evidence of cross-subsidization among risk classes, whereas the latter found no such result. Cohen (2005), with data from an insurer in Israel, found some evidence of adverse selection for drivers with more than three years of driving experience.

However, more recent studies have reported less clear results. For instance, focusing on automobile insurance, Chiappori and Salanié (2000) showed that drivers with comprehensive insurance do not have a statistically different accident frequency than drivers with minimum coverage, controlling for all observable characteristics. Richaudeau (1999) and Saito (2006) obtained similar conclusions.44 Dionne et al. (2001) verified that the results of Puelz and Snow (1994) could be explained by an improper econometric specification. They concluded that there is no residual asymmetric information on risk type in their data from an insurer in Canada, once the information provided by an efficient risk classification policy has been taken into account.

44 Other recent studies include those by Kim et al. (2009); Olivella and Vera-Hernández (2013); Dardanoni et al. (2018); and Geyer et al. (2020).
in the econometric specification. This led to fundamental questions being raised about the proper tests for adverse selection (see Cohen and Siegelman, 2010, for a review of empirical tests). More precisely, it seems that adverse selection plays a significant role in some insurance markets (annuities, crop insurance), but not in others (automobile, life insurance), and that the evidence is mixed for still other markets (health). This is due partly to the inability of individuals to use their information advantage. It is also due to the difficulty of dissociating adverse selection from moral hazard in claim observations when dynamic information is not available to researchers. Evidence of adverse selection requires that individuals with comprehensive insurance coverage report higher average claims than individuals with partial coverage or uninsured individuals. But higher average claims for fully insured individuals may also be due to moral hazard. For this reason, a positive correlation observed between risk and insurance coverage does not necessarily signal the presence of adverse selection, although a null correlation signals that adverse selection, as well as moral hazard, do not play a role. Moreover, as soon as empirical tests are considered, a simplifying assumption of the original Rothschild-Stiglitz (1976) model is enhanced, namely, of identical individuals, except for their accident probability. In particular, they all have the same degree of risk aversion. Of course, in reality, attitudes toward risk differ. As the degree of risk aversion is one factor influencing the demand for insurance coverage, it becomes difficult to conclude whether or not individuals who get more coverage from their insurers can be considered high risks. It may also happen that they belong to the good-risk group but demand more insurance simply because they are more risk averse (Smart, 2000; Villeneuve, 2003; Cohen and Einav, 2007; Chiappori and Salanié, 2013). The problem gets worse if, as Einav and Finkelstein (2011, p. 124) remark, “in many instances individuals who value insurance more may also take action to lower their expected

46 See Makki and Somwaru (2001).
cost: drive more carefully, invest in preventive health care, and so on.” Such a remark opens the door to the possibility of “advantageous selection”: the insureds with high degree of coverage are those with the lowest accident probability. They demand more insurance, even at a high price, not because they are high risks, but because they are, on average, more risk averse than the high risks, and the heterogeneity in risk-aversion coefficients exceeds the heterogeneity in endowed riskiness.  

Advantageous selection is not only a textbook curiosity. It has been documented in several markets, such as long-term care insurance (Finkelstein and McGarry, 2006) and supplemental insurance (Fang et al., 2008). In these markets, the correlation between observed risk and insurance coverage is negative rather than positive.

To overcome these pitfalls in testing for adverse selection, Einav et al. (2010) used identifying variations in the price of health insurance, provided by one specific insurer, to estimate the demand for insurance. The resulting variations in quantity, together with cost data, were then used to estimate the marginal cost of additional policies. This information allowed them to test for adverse selection (the marginal cost of contracts should be decreasing) and for the associated welfare loss. Their study provides evidence of adverse selection, but the welfare impact of this inefficiency seems to be small, in both absolute and relative terms.

When adverse selection is present, other insurance devices, used to deal with it, are experience rating and risk categorization. They may be used as substitutes or complements to separating contracts. Dionne (1983) and Dionne and Lasserre (1985) on one hand, and Cooper and Hayes (1987) on the other hand, extended Stiglitz’s (1977) monopoly model to multiperiod contracts, respectively, with an infinite horizon and a finite horizon, and with full commitment by the insurer to the terms of the contract. Hosios and Peters (1989) extended the finite-horizon case to limited commitment. In this case, contract renegotiation becomes relevant, as

---

49 Advantageous selection can lead to too much insurance being purchased if there are transaction costs and competition among insurers drives profits to zero. In equilibrium, the marginal cost of insurance exceeds the market price (see Einav and Finkelstein, 2011). The possibility of advantageous selection was first introduced by Hemenway (1990), who termed it “propitious” selection, and it was later analyzed by De Meza and Webb (2001).
information on the risk types increases over time. In addition, strategic use of accident underreporting becomes an issue.

Cooper and Hayes (1987) also extended the Rothschild-Stiglitz (1976) model to a two-period framework. They were able to demonstrate the beneficial effect of experience rating under full commitment by insurers, even when the insureds have the opportunity to switch to a different insurer in the second period (semicommitment). At equilibrium, the competitive insurer earns a profit on good risks in the first period, compensated for by a loss in the second period on those good risks who do not report an accident. This temporal profit pattern was termed *highballing* by D’Arcy and Doherty (1990). A different model, without any commitment, and assuming myopic behavior by insureds, was proposed by Kunreuther and Pauly (1985). The nonenforceability of contracts implies that sequences of one-period contracts are written. Insurers’ private information about their customers’ accident experience allows for negative expected profits in the first period and positive expected profits on the policies they renew in subsequent periods (*lowballing*). Later, Dionne and Doherty (1994) proposed a model assuming private information among insurers about the loss experience of their customers and semicommitment with renegotiation. Then, the insured has the option to renew their contract on prespecified conditions (future premiums are conditional on prior loss experience). This assumption seems to come closer to actual practices used in insurance markets. The authors derived an equilibrium with first-period semipooling and second-period separation. Their model predicts highballing, since a positive rent must be paid in the second period.

---

50 Note, however, that there exist alternative views on the welfare effect of asymmetric information. Using a two-period model where insureds have the option to switch insurers in the second period, de Garidel-Thoron (2005) showed that information sharing among insurers is welfare-decreasing. This is because it reduces the set of viable long-term contracts available to individuals in the first-period competition game.

51 In Kunreuther and Pauly (1985), the insurers have no information about other contracts their customers might write. For this reason, price-quantity contracts are unavailable. The equilibrium is a pooling equilibrium with partial insurance for the good risks, as in Pauly (1974).

52 In the first period, insureds may choose either a pooling contract with partial coverage and possible renegotiation in the second year, or the Rothschild-Stiglitz contract designed for high risks.
period to the high-risk individuals who experienced no loss in the first period, and this is compensated for by a positive expected profit on the pooling contract in the first period. Their empirical test, based on data from Californian automobile insurers, provides some support for this prediction. They conclude that some (but not all) insurers use semicommitment strategies to attract a portfolio of predominantly low-risk drivers.

More recently, Crocker and Snow (2011) drew attention to the fact that multidimensional screening is routinely used by insurers to cope with adverse selection. With $n$ mutually exclusive perils, insurers “can now exploit $n$ signalling dimensions to screen insurance applicants” (p. 293). The good risks tend to accept higher deductibles for perils they are less likely to be exposed to, for instance theft. This allows insurance markets to circumvent the nonexistence problem identified by Rothschild and Stiglitz (1976). As the authors themselves remarked, using multidimensional screening at a point in time presents an analogy with (and may be a substitute to) using repeated insurance contracts in a dynamic framework.\footnote{For good risks who do not file a claim in the first period, the reward takes the form of additional coverage in the second period.}

Risk categorization, which uses statistical information on correlations between risk classes and observable variables (such as age, sex, domicile, etc.), was first studied by Hoy (1982), Crocker and Snow (1986), and Rea (1992).\footnote{See also Bonato and Zweifel (2002) on the use of multiple risks to improve the assessment of loss probability.} Their work shows that risk categorization enhances efficiency when classification is costless but that its effect is ambiguous when statistical information is costly (Bond and Crocker, 1991). This last result was recently challenged by Rothschild (2011) who showed that a ban on risk categorization is always suboptimal—even when categorization is costly. He introduced a distinction between a regime where categorization is used by insurers (as in Crocker and Snow, 1986) and a regime where categorization is allowed, but may or may not be used in equilibrium. He then showed that a ban on risk categorization is always (regardless of the insurance market regime) Pareto-
dominated by having the government introduce a partial social insurance and simultaneously lifting the ban on risk categorization for private supplemental coverage.\(^{56}\)

Here is a quoting from Rothschild (2011, p. 269):

The intuition behind the effectiveness of social insurance for preventing the negative consequences of lifting categorical pricing bans is simple. Categorical pricing bans are potentially desirable insofar as they implicitly transfer resources from individuals in low-risk categories to individuals in high-risk categories. (...) Providing partial social insurance effectively socializes the provision of this cross-subsidy. Lifting a categorical pricing ban then allows the market to employ categorical information to improve efficiency without risking undoing the cross-subsidy.

These results are of the utmost political importance, given the ethical critiques on the use of observable personal attributes, such as sex and race, in insurance rating. The problem of risk categorization is even more acute when the personal attributes are not observable \textit{a priori} but may be revealed to the insurer and/or the insured after some information-providing steps have been taken, as in the case of genetic diseases. Rothschild and Stiglitz (1997) point out that this results in a conflict between the social value of insurance and competition among insurers. If valuable information can be made available about the probability (or certainty) that an insured will suffer from a particular genetic disease, then insurers will want that information. But this will result in less insurance coverage, since the insureds who are virtually certain to get the disease will not be able to obtain insurance, whereas those who are revealed to be immune to the disease will no longer need insurance.\(^{57}\) For ethical reasons, society prohibits the use of genetic information by insurers to categorize risks. But this means that adverse selection problems are enhanced, at least in medical insurance. As Doherty and Posey (1998) showed,

\(^{56}\) See Dionne and Rothschild (2014) for a framework to evaluate the economic consequences of legalized and banned risk classification systems, both in a static environment and a dynamic environment with learning.

\(^{57}\) This is an example of the well-known result that additional public information may have adverse welfare consequences (see, e.g., Arrow, 1978).
private testing is encouraged when test results are confidential and there is a treatment option available, but the insurers are unable to charge different prices to different customers with private information about their genetic background. Combining partial social insurance with supplemental private insurance, as suggested by Rothschild (2011), could be a way out of this conflict between the efficiency of insurance pricing and the mutuality principle.

Like moral hazard, adverse selection is an important problem that extends beyond the field of insurance. It is mainly encountered in labor markets, where employers are uninformed about the productivity of prospective employees, and in financial markets, where banks and finance companies lack information on the reimbursement prospects of borrowers. The insurance economics literature on adverse selection reviewed above has thus led to applications in other areas of economics: see, e.g. Miyazaki (1977) for an application to the labor market and Stiglitz and Weiss (1981) for an application to credit markets. Note, however, that in these cases, quality signalling by the informed agents represents a feasible strategy for circumventing the asymmetric information problem (Spence, 1973). For example, education and dividend payments find additional justification in these circumstances. By contrast, signaling does not generally occur in insurance markets, as insureds do not engage in specific activities to signal that they are good risks.

4.3. Moral hazard and adverse selection

Progress in analyzing moral hazard and adverse selection together has remained very limited. This was already noted by Arnott (1992) in the early nineties, and the situation has not significantly changed since then. This has long limited the significance of empirical investigations in the economics of insurance, since both problems come together in actual insurance markets. A positive correlation between insurance indemnities and insurance

---

58 In contrast, Doherty and Thistle (1996) found that additional private information has no value if there is no treatment option conditional on this information.
coverage may be interpreted as signaling the presence of adverse selection, or moral hazard, or both. The first attempts to address the two problems jointly were made by Dionne and Lasserre (1987) in the monopoly case, and by Eisen (1990) in the competitive case, but neither were echoed in the literature. In the same period, Bond and Crocker (1991) pointed out that risk categorization may be endogenous if it is based on information on consumption goods that are statistically correlated with an individual’s risk (correlative products). Thus, adverse selection and moral hazard become related in their framework. If individual consumption is not observable, then the taxation of correlative products by the government may be used to limit moral hazard, which would reduce the need for self-selection mechanisms as an instrument for dealing with adverse selection. However, they did not provide a general model.

Advances on this research front currently seem more promising at the empirical level, using material provided by longitudinal data on insurance purchasing and loss experience. Adverse selection is due to differences in the dynamics of learning about the insured’s true risk type for the insured themself and for the insurer. Once adverse selection has been identified (or not), then, with the use of longitudinal data, the residual effect of moral hazard may be tested. This is the approach followed by Dionne et al. (2013) using data on automobile insurance and car accidents in France. They first calibrated a simulation model for the optimal behavior of car owners, using the specific features of auto insurance in France, and showed that adverse selection and moral hazard should be expected. They then tested the model with real data for the presence of asymmetric learning on the one hand, and moral hazard on the other. Asymmetric learning is about accidents, not claims, and insurers observe only claims, while insureds observe both. In the long run, asymmetric learning can be shown to become adverse selection. The statistical results differ according to the experience of the insureds. For insureds with less than 15 years of experience, they found strong evidence of moral hazard but little evidence of asymmetric learning. The latter occurs only for insureds with less than five years of experience. By contrast, for insureds with more than 15 years of experience, there is no
evidence of moral hazard or adverse selection. These results should stimulate further research along the same line in other insurance contexts or markets.

5. Developments: Insurance market structure

Numerous studies on the insurance sector have followed the lead provided by Joskow (1973). The fact that most of these studies pertain to the US market can be explained by the availability of data and the size of the insurance market.

5.1. Insurance distribution

Insurance distribution systems have been analyzed by several researchers, more particularly Cummins and VanDerhei (1979) and Berger et al. (1997). In line with Joskow (1973), direct writing has generally been found to be more cost-efficient than independent agents. However, the differences in efficiency are not significant, which may be interpreted as an indication that independent agents provide valuable services. This interpretation has received support in several articles, e.g., Barrese et al. (1995) for the United States and Eckart and Räthke-Döppner (2010) for Germany. A comprehensive study on this topic is Cummins and Doherty (2006). The authors showed that insurance intermediaries (brokers and independent agents) have a valuable role in improving the efficiency of the market. Over the past 30 years, insurance markets in several European and Asian countries have witnessed the marketing of insurance contracts through banking channels—bancassurance. The efficiency of this distribution system has been investigated, with mixed results so far; for instance, Chang et al. (2011) do not report efficiency gains for this system in Taiwan.

Hilliard and Tennyson (2024) presented a detailed overview of insurance distribution systems in the US and international markets. They observed an increase in the standardization

---

59 See also Zweifel and Ghermi (1990) for a study using Swiss data.
of simpler insurance products in the life and nonlife markets over recent years. For these products, there is an emphasis on low-cost distribution and nontraditional methods of reaching clients, including the Internet. Customers pay their own premiums, and insurers use individual underwriting models, thereby significantly reducing distribution costs.

The industry is moving away from traditional relationships between insurer direct writing and independent agents. Professional agents are now more focused on selling more complex products such as commercial insurance. We observe fully integrated distribution systems without professional agents versus agency systems of distribution instead of the traditional separation between independent agents and a tied agency team. According to the authors, this can be explained by technological changes and increased competition in the insurance markets. The regulation of insurance distribution should now focus on the market conduct of professionals in the service-oriented sectors and on adequate disclosure in standardized insurance products.

5.2. Organizational structure

The various forms of organizational structure in the insurance industry—stock companies, mutuals, Lloyds’ underwriters—were analyzed in an agency theory framework by Mayers and Smith in a series of papers (1981, 1986, 1988), among others. They verified that conflicts of interest between owners, managers, and policyholders affect the choice of organizational form for different insurance branches (see also Hansmann, 1985; and Cummins et al., 1999b). Mutuals tend to prevail when the relationship between owners and policyholders triggers substantial agency costs. However, mutuals are constrained by their lack of access to external capital, with the result that mutuals with strong growth choose to convert to the stock structure when the constraint on their expansion becomes too costly: see also Mayers and Smith (2002), Harrington and Niehaus (2002), and MacMinn and Ren (2011) for other references. Ho
et al. (2013) showed that mutual insurers have a lower total risk, underwriting risk, and investment risk than do stock insurers. This research subject has not significantly evolved, per se, over the last ten years. However, organizational structure is often used as a control variable in empirical insurance analyses, such as corporate risk management, mergers and acquisitions, insurance pricing, and corporate governance.

5.3. Mergers and acquisitions (M&A)

Chamberlain and Tennyson (1998) showed that targets with greater capital needs are more likely to obtain a capital infusion at acquisition, particularly after 1985, when targets were poorly capitalized. Cummins et al. (1999a) showed that target life insurers experienced larger efficiency gains than did firms in a control group that were not involved in M&A deals. Cummins and Xie (2008) did a similar analysis in the property and liability sector for the 1994–2003 period. They obtained that targets achieve higher cost and allocative efficiency, as compared to acquirers. Being acquired is value-enhancing in their sample of firms. They also verified that poorly performing targets with low capitalization and poor underwriting performance are more likely to be targets, while large and rapidly profitable firms are more likely to be acquirers. Groups are more likely to be acquirers, compared to unaffiliated single firms and mutuals. Shim (2011) showed that the performance of acquiring firms decreases during the gestation period after the M&As, and that more focused insurers outperform product-diversified insurers.

Boubakri et al. (2008) investigated whether M&A transactions created value for acquirers’ shareholders during the 1995–2000 period, where acquirers are US property-liability insurers, and targets could be US or foreign insurers. They added a corporate governance dimension to their analysis by testing if the long-run performance of the acquirers is related to the percentage of shares held by institutional investors, blockholders, and the CEO, in the
absence of entrenchment problems. Their results confirm positive abnormal returns in the long run for acquirers. The returns are higher when buying other property-liability insurers, meaning that focused M&A transactions are more value-enhancing. Results related to corporate governance indicate that the percentage of shares held by the CEO and the CEO duality (CEO and board chair) are negatively related to the bidders’ long-run performance, which is consistent with entrenchment theory.

More recently, Cummins et al. (2015) tested the hypothesis that M&As can be value-creating for both acquirers and targets. They also tested whether cross-border transactions are value-creating for targets and acquirers. They verified that the transactions are value-enhancing for both acquirers and targets, with a greater effect for targets. Moreover, for acquirers, cross-border deals provide greater value, while there is no statistical difference for targets.

Klumpes (2022), with European data, showed that technical efficiency gains were significantly higher for acquiring insurers than for non-acquiring insurers in the pre-financial crisis period. Results also indicate that larger insurers, with a higher premium/surplus ratio, realized the highest efficiency gains. Well-capitalized insurers, with high ratios of equity / invested capital and capital / total assets realized a loss of efficiency for the sub-period of 1997–2007. Klumpes (2022) also estimated the probability that an insurer becomes an acquiring firm, using a dummy dependent variable taking the value of zero for non-acquiring insurers and 1 for acquiring insurers. Acquiring firms are included in the logistic regression analysis only in the year of their acquisition deal, and non-M&A firms are included for all sample years for the pre-crisis and post-crisis sub-periods. The logistic regression results show that more efficient insurers are more likely to be acquirers, which tends to support the previous result related to the technical efficiency improvements for M&A transactions. Results also indicate that log of assets, life insurance dummy, percentage of equity investments, and a UK
dummy are significantly positively related to the probability of being an acquirer in both the pre- and postcrisis periods.

5.4. Regulation

Following the lead provided by Joskow (1973), the effects of rate and solvency regulation have been scrutinized in a variety of research, such as that of Borch (1974), Ippolito (1979), Munch and Smallwood (1980), Danzon (1983), Finsinger and Pauly (1984), Pauly et al. (1986), Harrington (1984, 1987), Cummins and Harrington (1987), D’Arcy (1988), Harrington and Danzon (1994), and Cummins et al. (2001). These studies were stimulated by the traditional government regulation of insurance activities, a general trend toward deregulation in the eighties and nineties, and consumer pressures for re-regulation (mainly in California and Florida) after major disasters such as Hurricane Andrew in 1992 and the Northridge earthquake in 1994. Dionne and Harrington (1992) concluded their survey of research on insurance regulation by noting the following. Firstly, “not much is presently known about the magnitude of the effects of regulatory monitoring and guaranty funds on default risk” (p. 32). Secondly, rate regulation seems to have produced a variety of effects. It favored high-risk groups, increased market size and encouraged insurers’ exits, but nonetheless reduced the ratio of premiums to losses and operating expenses. More recently, Klein et al. (2002) found that price regulation tends to increase leverage, while Doherty and Phillips (2002) remarked that, during the nineties, when there was a trend toward deregulation, the role of rating agencies was enhanced: stringency in the rating procedures provided an incentive to decrease leverage and seemed to substitute for tight regulations. Rees et al. (1999), considering the focus of the European Commission on solvency regulation instead of rate regulation, suggested that “the

---

60 Berry-Stölzle and Born (2012) provide an empirical account of the deregulation introduced in Germany in 1994. They find evidence of a significant price decrease in highly competitive lines, offset by higher prices in the other lines.
role of regulation in insurance markets should be confined to providing customers with information about the default risk of insurers” (p. 55).

Debates on insurance regulation were exacerbated by the financial crisis in 2007–2008 and the doubts about insurers’ solvency following the collapse of Lehman Brothers and the rescue of AIG by the federal government. Eling and Schmeiser (2010) derived 10 consequences of the crisis for insurance supervision, while Lehmann and Hoffmann (2010) stressed the differences between insurance and banking. Harrington (2009) reviewed the AIG case and questioned the exposure of the insurance sector to systemic risk. He noted that this sector remained largely on the periphery of the crisis, in contrast to AIG (see also Kessler, 2014). He also noted that the crisis revealed the imperfect nature of federal regulation of banks and related institutions. These considerations led him to reject the plans for creating a federal systemic risk regulator for insurers and other nonbank institutions designated as systemically significant. He also rejected the claim that the AIG crisis strengthens arguments for federal insurance regulation—be it optional or mandatory. In his view, “an overriding goal of any regulatory changes in response to the AIG anomaly should be to avoid further extension of explicit or implicit ‘too big too fail’ policies beyond banking” (p. 815). Notwithstanding these strong arguments against federal involvement in US insurance regulation, the debate continued and the US federal government revisited the regulatory system in the McCarran-Ferguson Act. The 2010 Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank) gave increased systemic risk regulatory authority to the Federal Reserve. In addition, Dodd-Frank also created a Federal Insurance Office within the Department of the Treasury to establish greater uniformity among states with regard to insurance and reinsurance information reported to the federal government and to participate to international treaties (Grace, 2024).

61 AIG’s failure is mainly attributed to two causes. First, a subsidiary of AIG—AIG Financial Products—became heavily involved in the writing of credit default swaps. Second, another subsidiary, operating in the life branch, had engaged in securities-lending programs that were severely hurt by the onset of the subprime crisis. In either case, insurance operations were not concerned.
In a comprehensive review of insurance regulation procedures, Klein (2012)—see also Klein and Wang (2009)—spelled out the principles for insurance regulation and compared the traditional system of detailed state-by-state regulation in the US to the principles-based approach introduced in EU member countries under Solvency II. He concluded that, compared to the European developments, “the systems for solvency and market conduct regulation in the United States warrant significant improvement,” and that “the (US) states should move forward with full deregulation of insurance prices” (p. 175).

5.5. Insurance cycles

A related avenue of research, not considered by Joskow (1973), deals with cycles in the insurance industry. It was noticed in the seventies that insurance company profits seem to be subject to more or less regular cycles, and that this phenomenon is reflected in cyclical capacity and premium rates. The Geneva Association sponsored one of the first investigations in this area (Mormino, 1979). The most frequently quoted papers were published a few years later by Venezian (1985), Cummins and Outreville (1987), and Doherty and Kang (1988). As pointed out by Weiss (2007, p. 31), “in tracking underwriting cycles, most of the attention tends to be directed at insurance pricing, or, conversely, insurance underwriting profits, rather than the amount of coverage available.” The US insurance liability crisis of the mid-eighties and the overcapitalization of property-liability insurers during the nineties spurred research on insurance cycles (Harrington, 1988; Cummins and Doherty, 2002). Haley (1993) and Grace and Hotchkiss (1995) documented the impact of external factors—mainly interest rates—on underwriting profits, but the latter authors found that unanticipated external real economic shocks have little effect on underwriting performance. Other research has suggested that delays in adjusting premiums to expected claims costs—due to regulation or structural causes, external

62 The Solvency II regulation is presented and analyzed in Eling et al. (2007), Gatzert and Wesker (2012), and Doff (2024).
shocks to supply capacity, and variations in insurer insolvency risk—are responsible for
insurance pricing models as theories of the underwriting cycle. They showed that two models
are consistent with short-run and long-run data on underwriting profits: the capacity constraint
model and the actuarial pricing model. Cummins and Nini (2002) provided empirical evidence
that the overcapitalization of property-liability insurers during the nineties was mainly due to
capital gains during the stock market boom and to retained earnings. They interpreted their
results as providing evidence that insurers tend to hoard capital in favorable periods as a hedge
against future adverse shocks. This behavior negatively affects underwriting profits and is a
source of profitability cycles.

Following the lead provided by Cummins and Outreville (1987), research on insurance
cycles has also been conducted at the international level and has provided evidence that these
cycles are not specific to the US market: see Lamm-Tennant and Weiss (1997), Chen et al.
(1999), Leng and Meier (2006), and Meier and Outreville (2006). 63

Researchers became more skeptical about the existence of pricing and profitability cycles
in the property-liability insurance industry (Boyer et al., 2012; Henriet et al., 2016). One critical
factor was the methodology used in previous studies. Grace (2023) proposed a new
methodology to detect insurance profit cycles and used quarterly firm data, instead of yearly
industry aggregate data, to isolate firm-level heterogeneity. He found evidence that firms can
be in high- and low-loss ratio regimes by using a Markov switching model, but he did not
observe cyclical changes in pricing and profitability over time, confirming the results in the two
above contributions.

63 See Harrington et al. (2013) for a survey of the literature.
5.6. Insurance of catastrophes

The insurance of catastrophes has become a major concern for the industry and the subject of intensive academic research. Beyond the insurance industry, catastrophes have become an issue for the economy at large. Events like Hurricane Katrina on the Gulf Coast in 2005 or the Fukushima catastrophe in Japan in 2011, with its sequential earthquake, tsunami, and major nuclear accident, had repercussions for the international economy, not only locally. Given that major catastrophic events occur every year somewhere in our globalized world, it has become inappropriate to continue to define catastrophes as “Low-probability/High-consequences events” (see Kunreuther and Michel-Kerjan, 2009, p. 351). The major journals in the economics of insurance and some general economic journals have devoted special issues to this topic over the past two decades. Books and edited collections have also addressed this issue. Researchers have tended to take a broad view of the subject, so the term catastrophe has been used to encompass different kinds of situations: not only natural disasters (like earthquakes, tsunamis, floods, and hurricanes) and man-made catastrophes (such as Chernobyl or Bhopal), but also socioeconomic developments that result in a catastrophic accumulation of claims to insurers (Zeckhauser, 1995). The prominent example of this last category is the liability crisis in the US due to the adoption of strict producers’ liability and the evolution in court assessments of compensations to victims, as in the cases of asbestos, breast implants, pharmaceuticals, etc. (Viscusi, 1995). To cope with the financial consequences of catastrophes, traditional insurance and reinsurance are often considered insufficient (Kunreuther, 1996; Froot, 1999a, 1999b; Cummins et al., 2002). Some researchers have invoked the difficulties individuals would have in dealing with low-probability/high-loss events (Kunreuther and Pauly, 2004). Others have invoked capital market imperfections and market failure in reinsurance supply (Froot, 2001; Zanjani, 2002; Froot and O’Connell, 2008). Still others have pointed to US insurance price

---

regulation in catastrophe-prone lines of business as a major source of inefficiency in the insurance and reinsurance markets (Cummins, 2007). Several researchers have advocated for more government involvement (Lewis and Murdock, 1996; and Kunreuther and Pauly, 2006), but others have argued that the government has no comparative advantage over the market in providing coverage for catastrophic losses (Priest, 1996) and have instead called for less government intervention by deregulating insurance prices (Cummins, 2007). Alternative solutions may be found in financial innovation, either in the design of insurance contracts, by introducing a decomposition of insurance risk into a systemic and a diversifiable component (Doherty and Dionne, 1993; Schlesinger, 1999; and Doherty and Schlesinger, 2002), or in the design of new financial securities, or both.

5.7. Climate finance

Climate finance is defined by the United Nations Framework Convention on Climate Change (UNFCCC) as being “local, national, or transnational financing—drawn from public, private, and alternative sources of financing—that seeks to support mitigation and adaptation actions that will address climate change” (Reported in Harrison et al., 2020). This concerns the financing that must be undertaken to change the world economy and build resilience to climate change.

Many financial sectors (banking, insurance, real estate) are directly impacted by the risks generated by carbon prices, tornadoes, wildfires, pandemics, and flooding. Difficult questions are raised: How should financial markets price and mitigate risks from global warming? How can capital markets raise so much financing? How should the distribution of damages from catastrophic events be managed? These were recently discussed in a special issue of the Review of Economic Studies, edited by Harrison et al., (2020). But there are no contributions available

---

65 Monti (2011) provides a recent review of public-private arrangements that already exist in the OECD area.
in finance and insurance journals on the causal effects of climate change on the insurance industry.

A crucial input for the analysis of climate change risks is the causal impact of climate events on economic activity, what is called the distribution of damages. This raises an important question about the modeling and sharing of extreme weather risks. Do extreme weather risks, such as the impact of Hurricane Sandy in 2012 or the 2018 California wildfires, have long-run causal effects on insurance markets? These distributions of damages depend on locational decisions by households and firms and on technology (self-protection and self-insurance) decisions, in terms of preventing and mitigating the damage of disasters. They also depend on market insurance coverage (including moral-hazard and adverse-selection effects). In suitably modeling these loss distributions, the insurance industry should be able to play a critical role in facilitating risk-sharing and extending insurance coverage for extreme weather events. In addition, public authorities must improve their prevention role.

The potential impacts of new climate patterns on damages from catastrophic risks must be better estimated by the insurance industry and public authorities. These potential impacts may have been underestimated in risk management for many years. Here are some worldwide statistics according to a Munich Re report (2019):

- 83% of all natural disasters worldwide were weather-related during the 1980-2019 period; and 43% of the overall economic losses in that period occurred in Asia.
- 64% of the insured losses were incurred in North America (incl. Central America and the Caribbean) during this period, which represents about 30% of overall losses in this region and around the world.
- According to the World Meteorological Organization, 11,000 climatic events were observed from 1970 to 2019, for a total cost of US$1,400B.
In 2008, the worldwide economic losses from climate disasters was evaluated at $130B, of which $44B was insured. Climate risk was rated number one of the top 10 risks facing the insurance sector (Ernst and Young, 2008). The average total loss over the last 10 years is $187B ($340B in 2017 only). The year 2019 was below the last 10-year average, with a total loss of $150B but with 33 events with more than 1B in total losses each. Nine events cost the insurance industry more than $1B that year, and all of them were climate risk events (cyclones, storms with flooding, and tornadoes). And this is not the end. In April 2020, severe weather events in the US cost insurers billions of dollars, which had 14 tornadoes—the fifth worst month on record since 1950, according to an Aon report (2023).

The escalating frequency and severity of extreme weather events highlight a dangerous link between insurance risk and climate change, even though less than 40% of total losses are covered. According to a recent survey by Deloitte (2020), most US state insurance regulators expect all types of insurance companies’ climate change risks to increase over the medium-to-long term. More than half the state regulators surveyed also indicated that climate change is likely to have a high impact on coverage availability and underwriting assumptions. US state regulators and lawmakers are concerned about the insurance industry’s response to climate change. Two traditional mechanisms are usually used to reduce financial fragility. Insurers can increase premiums in states or counties most affected, or they can increase reinsurance coverage (Grenier, 2019). But these two alternatives may not be satisfactory for the long-run stability of the industry, even though they have been used by the US property and casualty insurance industry over the last 30 years (Dionne and Desjardins, 2022).

Insurance-linked securities (ILS) can lower the cost of risk transfer in hard (re)insurance market conditions. They help maintain the (re)insurance capacity and offer multiyear protection. They are free of credit risk by offering full collateralization of losses. For investors, they are noncorrelated with other market, liquidity, and credit risks, so they represent an
important diversification asset. Alternative capacity (ILS) includes collateral reinsurance, sidecar, Industry Loss Warranty (ILW), and CAT bonds. ILS penetration can reduce the price of insurance in the long run and increase the demand for insurance. But there is a long-run commitment issue as regards the financial participation of financial markets.

According to data in 2019, the total international reinsurance capital was about $625B, including $62B in ILS capacity. It should be mentioned that the total capital of the US insurance industry was about $1.1 trillion in 2020 (Dionne and Desjardins, 2022). In complement to reinsurance, ILS accounted for only about 10% of global catastrophe reinsurance capital in 2019. We might think there is sufficient capacity, because annual average long-run losses are below $150B, but there have been significant recent exceptions: 2011 ($160B), 2017 ($170B), with an annual growth of 5–7%.

Mills (2009) analyzed different mechanisms to improve the insurance industry's capacity to cover insurable losses: new coverage products, a better understanding of climate changes, and the financing of activities designed to reduce climate risk. Gollier (2005) insisted on the government playing a necessary role in reducing the fragility of the insurance industry when there are extreme events. The government should act as a reinsurer to reduce bankruptcies. This complements Mills (2009), who favored stronger private risk management. (See also Michel-Kerjan, 2012, and Kunreuther, 2018).

Jametti and Ungern-Sternberg (2010) did not consider the observed risk selection between the private and public sectors to be optimal where the private sector keeps the acceptable or lower losses, and the public sector is limited to the extreme losses. Born and Viscusi (2006) proposed a different approach to analyze the effect of natural disasters on the insurance industry. Using data from the Swiss Re Sigma Reports for the 1984–2004 period, they showed that small insurers are more likely to be affected, because they are less diversified. Finally, Born and Klimaszewski-Blettner (2013) verified that some insurers tend to reduce their
activities when they are subject to severe regulations or to unanticipated large claims. This behavior is less frequent for large insurers that are better diversified but has become more common in recent years.

In their theoretical analysis, Cummins et al. (2002) studied the capacity of the US insurance industry to cover very large catastrophes. They proposed a sufficient condition for capacity maximization: all insurers must hold a net of reinsurance underwriting portfolios that are perfectly correlated with aggregate industry losses (Borch, 1962, model). Estimating capacity using insurers’ financial statement data, they found that the industry could adequately fund a $100B event, with data covering the 1983–1997 period, where the US insurers’ equity capital was approximately equal to $350B in 2002. Cummins et al. (2002) also showed that the industry would be able to pay very high percentages of the losses. Using their parameter estimates, the US insurance industry would have been able to pay about 81% of a $200B loss in 1997. Dionne and Desjardins (2022) showed that the capacity had increased to 98% in 2020.

5.8. Microinsurance

At the other end of the insurability spectrum, microinsurance has emerged as a new topic for research in insurance economics. The fact that a large fraction of the world’s population has no access to the benefits of insurance coverage has motivated practical initiatives to remedy this situation, as well as interest among researchers and international organizations, for instance the International Labour Office (ILO). A recent study (Biener and Eling, 2012) reviewed the current situation of microinsurance. The authors pointed out that the microinsurance industry has experienced strong growth in recent years (10% on average), but that much remains to be done and that further developments are constrained by well-known insurability problems: risk assessment, asymmetrical information, and lack of financial resources in the uninsured population. They also provided tentative solutions to overcome these problems. In a more recent
analysis of microinsurance markets, Dror and Eling (2024) concluded that the offer of low-cost insurance to these populations is still not effective.

5.9. Corporate governance

Corporate governance issues in the insurance industry have attracted more attention from researchers in the wake of the 2008 financial crisis and the collapse of AIG. The impact of corporate governance and institutional ownership on efficiency (Huang et al., 2011), risk-taking (Cheng et al., 2011; Ho et al., 2013), M&As (Boubakri et al., 2008), and CEO turnover (He and Sommer, 2011) have been investigated, among other issues.66

New risk regulations were introduced after the 2008 financial crisis. Boards of large insurers had to establish a risk committee, consisting of only independent members, and with at least one member having a thorough understanding of risk management. The role of the committee is to monitor the company’s risk appetite and oversee the risks the insurer is exposed to and the models used to measure, monitor, and mitigate them (Magee et al., 2019; Ames et al., 2018).

6. New approaches: Finance and insurance

Apart from the tremendous developments summarized in the three preceding sections, risk and insurance economics experienced a major refocus in the 1970s and 1980s, in which insurance was increasingly analyzed within the general framework of financial theory. This change of perspective was implicit in the definition of Arrow (1970): “insurance is an exchange of money for money.” It was also foreshadowed by the recognition that insurers are financial intermediaries (Gurley and Shaw, 1960). It soon became impossible to maintain the dichotomy

66 See Boubakri (2011), Boubakri and Valéry (2024), Anderloni et al. (2020), and the September 2011 Special Issue of The Journal of Risk and Insurance for recent surveys of corporate governance in the insurance industry.
in the analysis of insurance firms, with insurance operations on the one hand and financial investment on the other. As a result, insurance research became deeply influenced by advances in finance theory. This was heightened by the fact that finance underwent a major revolution in the 1970s, with the development of option theory, and that this revolution stressed the similarity between insurance products and new concepts coming from financial innovations (e.g., portfolio insurance).  

6.1. Portfolio theory and the CAPM

The influence of portfolio theory on the analysis of insurance demand was mentioned in Section 3. But this theory also had a profound influence on the theory of insurance supply. It was soon recognized that financial intermediaries could be analyzed as a joint portfolio of assets and liabilities (Michaelsen and Goshay, 1967), and this global approach was applied to insurance company management. Under this view, insurers have to manage a portfolio of correlated insurance liabilities and investment assets, taking into account balance-sheet and solvency constraints, and there is no justification for separating the operations into two separate areas: what matters is the overall return on equity (see Kahane and Nye, 1975, and Kahane, 1977).  

This way of looking at insurance operations led to a theory of insurance rating, reflecting the move observed a decade earlier in finance, from portfolio theory to the CAPM. Applying this model to insurance, it turns out that equilibrium insurance prices will reflect the undiversifiable risk of insurance operations. If insurance risks are statistically uncorrelated with financial market risk, then equilibrium insurance prices are given by the present value of expected claims costs (in the absence of transaction costs). If they are statistically correlated, a

---

67 The similarity between option contracts and insurance policies was stressed by Briys and Loubergé (1983).
68 See also Loubergé (1983) for an application to international reinsurance operations, taking foreign exchange risk into account, and MacMinn and Witt (1987) for a related model.
positive or negative loading is observed in equilibrium. The model was developed by Biger and Kahane (1978), Hill (1979), and Fairley (1979). It was empirically evaluated by Cummins and Harrington (1985). It was also applied to determine the “fair” regulation of insurance rating in Massachusetts (Hill and Modigliani, 1986).

In their recent review of the literature, Bauer et al. (2024) discussed in detail the difficulties of reconciling theory and practice in insurance pricing models. Particular attention was given to the CAPM model. First, the CAPM model is a one-period model, while insurance cash flows cover many years in many lines of business. A partial solution proposed by Myers and Cohn (1987) is to use the discounted net present value of loss payments. This solution involves estimating a liability beta to obtain a discount rate. This introduces a new difficulty, as only accounting information is available for many insurance liabilities.

Another solution is to use the cost of capital measured by the internal rate of return (IRR) approach. As shown by Bauer et al. (2023), the IRR model can be reconciled with the CAPM model when the discount rates of assets, liabilities, and equity are comparable. In fact, there is no need to compute the discount rate of liabilities if we know the two other rates. However, the IRR approach is not devoid of technical problems. Bauer et al. (2024) discussed other approaches, such as the risk-adjusted discounted cash flow model and present value of income over present value of equity. In conclusion, they mentioned that it is still difficult to divide insurance policy premiums between general market risk changes from financial markets, and frictional costs associated with holding capital.

---

69 Myers and Cohn (1987) extended the model to multiperiod cash flows, while Kraus and Ross (1982) considered the application to insurance of the more general arbitrage pricing theory.
6.2. Option pricing theory

Another limitation of the capital asset pricing model is that it does not take into account nonlinearities arising from features such as limited liability and asymmetric tax schedules. These aspects are best analyzed using option pricing theory, since it is well known that optional clauses imply nonlinearities in portfolio returns. Doherty and Garven (1986) and Cummins (1988) analyzed the influence of limited liability and default risk on insurance prices, while Garven and Loubergé (1996) studied the effects of asymmetric taxes on equilibrium insurance prices and reinsurance trade among risk-neutral insurers. A major implication of these studies is that loaded premiums not only reflect transaction costs and asymmetric information, or insurers’ risk aversion, they also reflect undiversifiable risk arising from institutional features, and they lead to prices implying risk-sharing in equilibrium, even when market participants are risk neutral.

The importance of option theory for the economics of insurance has also been observed in the area of life insurance. This resulted from the fact that competition between insurers and bankers, in order to attract savings, has led to the inclusion of numerous optional features (hidden options) in life insurance contracts. Advances in option theory have thus often been used to value life insurance contracts (see, e.g. Brennan and Schwartz, 1976; Ekern and Persson, 1996; and Nielsen and Sandmann, 1996) or to assess the effects of life insurance regulation (Briys and de Varenne, 1994). It seems, however, that the option pricing of insurance has not attracted much research over the last 20 years.

6.3. Insurance and corporate finance

The portfolio approach to insurance demand led to a paradox when applied to corporations, because these are owned by stockholders who are able to diversify risks in a stock portfolio. If insurance risks, such as accident and fire, are diversifiable in the economy, the
approach leads to the conclusion that corporations should not bother to insure for them. They would increase shareholders’ wealth by remaining uninsured instead of paying loaded premiums (Mayers and Smith, 1982). The paradox was solved using the modern theory of corporate finance, where the firm is considered a nexus of contracts between various players: managers, employees, suppliers, bondholders, banks, stockholders, consumers, etc. The reduction of contracting and bankruptcy costs provides an incentive to manage risk and to purchase insurance, even if the premium is loaded and shareholders are indifferent to insurance risk: see Main (1982), Mayers and Smith (1982, 1990), and Stulz (1984). In addition, increasing the marginal cost of external financing and convex tax schedules arising from progressive tax rates and incomplete loss offset offer other explanations for concern with insurance risk management in widely-held corporations: see Froot et al. (1993), Smith and Stulz (1985), Smith et al. (1990), Tufano (1996), Stulz (1996).

These considerations have changed the relationship of corporate managers to risk management in general, and to insurance in particular. These tools are no longer used simply because risks arise. They must find a justification in the firm’s overall objective of value maximization. Following these premises, several studies have addressed the relationship between corporate risk management and the capital-structure decision or the dividend policy. For instance, Auñon-Nerin and Ehling (2008) found that higher leverage increases the demand for corporate insurance and the use of derivatives, while hedging generally has a significant positive effect on leverage (see also Zou and Adams, 2008). This is in line with corporate concern for default and agency costs. They also found that corporate hedging is negatively related to the dividend payout ratio. This is related to the use of cash flows as a possible substitute for insurance (on this aspect, see also Rochet and Villeneuve, 2011). In addition, as

---

70 The same kind of argument was used by Doherty and Tinic (1981) to question the motivation of reinsurance demand by insurers.

71 Rochet and Villeneuve (2011) showed that cash-poor firms should hedge using financial derivatives but not insure, whereas the opposite is true for cash-rich firms.
Doherty (1997) noted, the development of financial engineering in the eighties challenged traditional insurance strategies in corporate risk management. Traditional insurance strategies often involve large transaction costs, and they fail if the risk is not diversifiable, as in the case of the US liability crisis. For this reason, innovative financial procedures, such as finite risk plans and financial reinsurance, represent alternative instruments for dealing with corporate risks. Of course, they also widen the competitive interface between banks and insurers.

The theory of corporate finance was also used by Garven (1987) to study the capital-structure decision of insurance firms. His paper shows that redundant tax shields, default risk, bankruptcy costs, and the abovementioned agency costs influence the insurer’s capital-structure decision. Plantin (2006) emphasized that reinsurance purchases and capital-structure decisions are linked. Professional reinsurers are used by insurers as a signal of credible monitoring sent to the financial market. Reinsurance is not only used as a device to mutualize risks, as in Borch (1962), or to address agency problems, as in Mayers and Smith (1990). It is also used as a complement in the insurer’s capital-structure strategy.

At the empirical level, Garven and Lamm-Tenant (2003), as well as Powell and Sommer (2007), provided evidence that reinsurance purchases are positively related to insurer leverage. Shiu (2011) used data from the UK nonlife insurance industry to test the two-way relationship between reinsurance and insurers’ capital structure. His results show that leverage exerts a positive influence on reinsurance purchases and that higher leverage is associated with more reinsurance purchases. He also found that the use of financial derivatives by insurers has a moderating impact on this two-way relationship.

More recently, Desjardins et al. (2022) studied the reciprocal causal relationship between reinsurance demand and liquidity creation by insurers. Financial institutions improve economic growth by creating liquidity in the economy or by investing in various assets or projects. Their generalized method of moments (GMM) estimation results show positive bicausal effects
between the two activities for small insurers, but no significant effects for large insurers, which do not seem to use reinsurance to protect their liquidity risk associated to long-term investments.

Taking a more general view, Hoyt and Liebenberg (2011) investigated whether an integrated risk management approach, as defined by Meulbroek (2002), has a positive influence on insurers’ value. They used Tobin’s Q as a measure of firm value and a maximum-likelihood procedure to estimate joint equations for the determinants of an integrated risk management policy and its impact on Tobin’s Q, for a sample of 117 publicly-traded US insurers. They found that insurers engaged in integrated risk management “are valued roughly 20 percent higher than other insurers after controlling for other value determinants and potential endogeneity bias” (p. 810). They also found that these insurers are larger, have less leverage, and rely less on reinsurance than other insurers.72

This result, added to those obtained above by Shiu (2011), can be related to earlier considerations by Doherty (1997) that insurers’ management has been deeply influenced by developments in the financial markets. The concept of asset-liability management, which has its roots in the portfolio approach mentioned above, means that insurers are less reliant on reinsurance as the natural instrument to hedge their risks and send signals to their partners. Indeed, developments in the financial markets over the past 20 years have resulted in the emergence of derivative products intended to complement traditional reinsurance treaties in an integrated risk management view (Cummins and Weiss, 2009).

6.4. Insurance and financial markets

In 1973, the insurance–banking interface was a sensitive subject. It was generally not well-considered, in the insurance industry, to state that insurance is a financial claim and that

72 See Liebenberg and Hoyt (2024) for a recent review of this literature, and Cummins et al. (2009) for the effect of risk management on the efficiency of insurance firms.
insurers and bankers perform related functions in the economy. Some 50 years later, and after numerous experiences of mergers and agreements between banks and insurers, the question is not whether the two activities are closely related, but where they differ.

It is easy for an economist of risk and insurance to provide a general answer to this question, founded on Borch’s mutuality principle (see Section 2) and on subsequent work on risk sharing. Insurance and banking, like all financial activities, are concerned with the transfer of money across the two-dimensional space of time and states of nature. Insurance deals mainly—but not exclusively (see life insurance)—with transfers across states that do not necessarily involve a change in social wealth. By contrast, banking and financial markets perform transfers across states that often involve a change in social wealth. In other words, insurance is mainly concerned with diversifiable risk; banks and finance companies (such as mutual funds and hedge funds) are mainly concerned with undiversifiable (social) risk.

This kind of distinction has been used before to draw a line between private and public (social) insurance. According to this view, social insurance is called for when the limits of private insurability are reached, in the sense that the insured events are not independent, so that diversifiability does not obtain: epidemics, losses from natural disasters, unemployment, etc. But, social insurance is limited by national boundaries. In absence of redistributive concerns and with market incompleteness due to asymmetric information, it has become increasingly obvious that financial markets are able to perform some social insurance functions, in addition to their traditional function of sharing production risk.

---

73 The convergence between reinsurance and investment banking was emphasized by Cummins (2005).
74 The debate has regained importance after the 2008 financial market crisis and the collapse of AIG. Large insurance companies have been ranked with banks in the group of “Systemic Important Financial Institutions” (SIFI) and are threatened to become subject to the same regulations as banks. This is an occasion for the insurance industry to underscore the differences between banking and insurance (see Lehmann and Hoffmann, 2010; and Geneva Association, 2010).
75 Public insurance may also be justified by equity considerations, e.g., in medical insurance.
A case in point is the evolution in the natural catastrophes branch of insurance. Since losses from natural disasters are correlated, they should be excluded from the private insurance area. Nonetheless, private insurance companies used to cover this risk because geographical dispersion seemed possible using the international reinsurance market. However, over the last two or three decades, the private insurability of this risk has been challenged by various developments such as the increased frequency and severity of hurricanes, huge losses, and a concentration of insured values in selected exposed areas of the globe: the US (mainly California and Florida), Japan, and Western Europe (mainly in the south). As a result, potential losses may have affected the financial capacity of the catastrophe reinsurance market (see Kielholz and Durrer, 1997; Cummins et al., 2002). On the other hand, Dionne and Desjardins (2022) pointed out that the US insurance market’s capacity has evolved over recent years. They showed that the capacity to pay for catastrophe losses is higher in 2020 than in 1997. This is explained by an increase in reinsurance demand and by higher premium rates for that risk. It is fair to say, however, that this increased capacity may not persist in the long run. Reinsurance is under pressure due to global warming and its observed impact on the severity and frequency of natural disasters around the world. Global warming reduces reinsurers’ traditional role of international diversification. It calls for the development of new tools able to cope with climate risk. ILS were considered an appropriate tool to complement reinsurance but their international outstanding volume, at about $40B, is still very far from the needs expected for future years (Artemis website).\(^7^6\)

A first possible solution to the insurability problem is the traditional recourse to government insurance using increased taxation, i.e., social insurance. This is the solution adopted in France (Magnan, 1995) and in some other countries,\(^7^7\) where a reserve fund financed

\(^7^6\) Adding collateralized reinsurance, ILWs, sidecars, and other financial coverage products, the outstanding international volume is at about $60B (Artemis website).

\(^7^7\) In the US, where a National Flood Insurance program has already existed for long, and where California has established a government earthquake insurance program (the California Earthquake Authority), the possible
by specific taxes on property-liability insurance contracts indemnifies victims from natural disasters. The viability of this solution is however endangered in the long run by increasing risk due to global warming and by wrong incentives: development of activities and construction in areas exposed to catastrophe risk, and pressure on the government to widen the scope of coverage while maintaining low rates.

A second solution is transferring the risk using special-purpose derivative markets. This was the solution proposed by the Chicago Board of Trade (CBOT) with the catastrophe options and futures contracts launched in December 1992: see D’Arcy and France (1992), Cummins and Geman (1995), and Aase (1999) for an analysis of these contracts. However, the CBOT contracts were withdrawn after some years due to lack of success. Following Hurricane Katrina in 2005, new contracts of the same type were nevertheless launched in 2007 by the Chicago Mercantile Exchange (CME) and the Insurance Futures Exchange. A main difference from the earlier CBOT contracts was their focus on US hurricanes and US tropical winds. Given the experience with the CBOT contracts, experts had doubts about the ultimate success of this new venture (see Cummins, 2012). But the New York Mercantile Exchange of the CME still provides such options and futures derivative contract, under the group of “weather contracts.” The geographical coverage was extended to Europe and Japan, even if the volumes are close to

---

78 The early options and futures on four narrow-based indices of natural catastrophes were replaced in October 1995 by call spreads on nine broad-based indices. Lewis and Murdock (1996) proposed to have the same kind of contract supplied by federal authorities in order to complete the reinsurance market.

79 Harrington and Niehaus (1999) had reached the conclusion that basis risk would not be a significant problem for Property Claims Services (PCS) derivative contracts, but later on, Cummins et al. (2004) reached a different conclusion, attributing the lack of success to basis risk. One might add that, possibly, the failure was due to the absence of arbitrage trading. Arbitrage trading between a derivative market and the market for the underlying instrument is essential to the provision of liquidity in derivatives trading for hedging and speculation purposes. However, in the case of PCS option contracts, such trading was impossible. The only market for the trading of insurance portfolios is the reinsurance market, which is not liquid enough for use as a vehicle in arbitrage trading.
zero, meaning that these contracts are unlikely to substantially complement reinsurance capacity in the future.  

A third solution is the securitization of the risk, using more familiar securities, such as coupon bonds, issued by a special-purpose vehicle (on behalf of an insurer, reinsurer, or nonfinancial company) or by a public agency (on behalf of the State): see Litzenberger et al. (1996) and Loubergé et al. (1999) for early presentations and analyses of insurance-linked bonds (widely known as CAT bonds). In a CAT bond arrangement, a special-purpose reinsurer (SPR) issues a coupon bond on behalf of the sponsoring entity and improves the return on the bonds with a premium paid by the sponsor. The principal is then invested into first-class securities, such as government bonds. However, on the investor side, the principal and the coupons are at risk, in the sense that they may be lost partially or even totally if a catastrophe occurs and the CAT bond is triggered. In such a case, the proceeds of the investment are used by the SPR to pay indemnities to the sponsor. The catastrophe risk has been transferred to the financial market using familiar securities as a transfer vehicle. In contrast to CBOT derivatives, these ILSs were well-received by the market. Their success was based on the huge pool of financial capacity provided by worldwide capital markets and the prospect of risk diversification made available to investors: catastrophic insurance losses are, in principle, uncorrelated with financial market returns. In addition, CAT bonds that have been based on an index of losses due to a specific catastrophic event, or triggered by such a specific event (parametric trigger), have made it possible to avoid the moral hazard arising from products based on the record of losses experienced by the sponsor.  

On the supply side, CAT bonds provide sponsors with coverage that extends over several years, at fixed terms (unlike reinsurance contracts), and free from default risk since the proceeds from the bond issue are

---

80 The pricing of these contracts was addressed by Perrakis and Boloorforoosh (2018).
81 Exposure to moral hazard for the investor is traded against basis risk for the sponsor.
fully collateralized using highly-rated securities.\textsuperscript{82} CAT bonds have attracted broad interest among insurance practitioners (see Swiss Re, 2009)\textsuperscript{83} and academic researchers (see Barrieu and El Karoui, 2002; Lee and Yu, 2002; Nell and Richter, 2004; Cummins, 2008; Michel-Kerjan and Morlaye, 2008; Barrieu and Loubergé, 2009; and Finken and Laux, 2009).\textsuperscript{84} The development of the market for these securities indicates that they filled a gap in the reinsurance market, although the success has not been as huge as initially anticipated: CAT bond and ILS issues started with a few issues prior to 2000, then the market peaked with 31 issues in 2007, and regained momentum in 2010 (26 issues) after a drop to 14 issues in 2008 due to the financial crisis. In 2022, 73 deals were observed, for a value of $10.5B (Artemis). The market is nevertheless developing over time, with positive and negative shocks being triggered by natural catastrophes and financial crises.\textsuperscript{85}

A fourth solution available to an insurer to hedge catastrophe risk outside the reinsurance market is provided by Catastrophic Equity Puts (CatEPuts). Under this arrangement, the insurer or reinsurer purchases from the option writer the right to issue preferred stocks at a specific price following the occurrence of a catastrophe. This allows the insurer to take advantage of fresh funding at a predetermined cost in a situation where recourse to the capital market would be prohibitive for them. It illustrates the increased integration of insurance and investment

\textsuperscript{82} The risk of default by the reinsurance provider is a concern in the high-layer segment of the reinsurance market.

\textsuperscript{83} Still, it remains that the use of insurance-linked securities raises sensitive issues in terms of regulation. Not because these instruments would represent a danger for the stability of the financial system, but because regulators, more particularly in the US, are reluctant to consider them as genuine alternative mechanisms for risk transfer: see Klein and Wang (2009).

\textsuperscript{84} The success with CAT bonds aroused interest for other insurance-linked securities, particularly in the life insurance sector (mortality bonds, longevity bonds): see Cowley and Cummins (2005); Lin and Cox (2005); Albertini and Barrieu (2009); Cummins and Weiss (2009); and Chen and Cox (2009).

\textsuperscript{85} See Cummins (2012) for a comprehensive report on the state of the market at year-end 2011, and the recent survey by Barrieu et al. (2024).
banking, both activities performing a fundamental economic function, i.e., the transfer of risks (Arnone et al., 2021).

7. Conclusion

In the early seventies, it was not clear how risk and insurance economics would develop over the years to come. Some 50 years later, it is comforting to realize that considerable developments have taken place: the length of the reference list below, unconventionally divided into pre-1973 and post-1973 references, gives an account of the qualitative and quantitative aspects of these developments.

This chapter focuses on developments that have mainly taken place along three avenues of research:

1. The theory of risk-taking behavior in the presence of multiple risks, which encompasses the theory of optimal insurance coverage, the theory of optimal portfolio investment, and the theory of optimal risk prevention. The dynamic behavior of economic agents over multiple periods is also considered.

2. The issues raised by asymmetric information for contract design and market equilibrium, a theme which extends beyond insurance economics and concerns all contractual relations in the economy, e.g., labor markets, product markets, and financial markets. Insurance economists have been very active in developing original empirical tests on the significance of asymmetric information, tests that can be applied in different markets.

3. The applications of new financial paradigms, such as contingent claims analysis, to the analysis of insurance firms, insurance markets, and corporate risk management, a

86 Other innovations, such as sidecars and ILWs, are different in nature from those presented in this section. They are innovations that improve the capacity of the reinsurance market, without introducing an alternative or complement to reinsurance contracts.
development that links insurance economics more closely to financial economics, and insurance to finance.

Nowadays, risk and insurance economics represent a major theme in general economic theory. Risk and insurance education, per se, has also become a predominant theme in many countries. But risk and insurance issues are still pervasive in economic education, more particularly in microeconomics. To verify this statement, one need only look at the post-1973 section of the following reference list to see that many important papers for the advancement of risk and insurance theory were published in general economic and financial journals, and not only in the leading specialized reviews. Indeed, given that this goal of the seventies was reached, one might wonder whether a different objective, that is, the development of more specialized risk and insurance education and research, which was given less importance back then, should not be reevaluated today.

Given the tremendous amount of research activity we have witnessed in the study of financial markets over the past years, we could infer that specialized research in insurance economics would receive a major impulse from the creation of complete, reliable, and easily accessible insurance databases. Compared to the situation at the end of the nineties, the last 30 years have been characterized by a breakthrough of empirical research in insurance economics, where the implications of models have been subject to empirical tests. These tests represent fundamental progress in the economics of risk and insurance. They provide results that enhance our understanding of insurance markets, and the authors must be congratulated for their efforts. But these contributions are still too often based on proprietary data, made available on a case-by-case basis, and not on widely available insurance databases, which reduces the power of the conclusions. Insurers are still reluctant to share their data. The availability of more databases would certainly trigger more interest in dissertations on risk and insurance from beginning
doctoral students in economics and finance. The development of strong causality tests must be based on high-quality data (Dionne, 2024).

Over the last 10 years, many new developments in the insurance literature have been observed. The solvency regulation of insurance firms became an important activity after the 2007–2009 financial crisis. Insurers must now be more transparent about their risk management activities, transmitting their yearly risk appetite statement to the regulator and having more board discussions about the computation of economic and solvency regulated capital. This computation must now be based on market data instead of only actuarial data. Better dynamic stress tests of capital must also be implemented.

Natural catastrophe risk has been omnipresent in insurer losses in recent years. Some insurers have left the market, and important questions about the international coverage of these losses via reinsurance have been raised because global warming limits international diversification. The development of alternative financial coverage is positive but not very significant, at about ten percent of the total reinsurance capital. Economic inflation is not well integrated in the management of insurance firms, along with social inflation costs. Cyber-risk costs and pandemic costs are becoming new sources of sizable losses. It is not clear that the insurers’ traditional risk management tools are effective for limiting these large losses and expenses, which are often correlated.

The insurance market faces many new innovations that affect the management of insurance and reinsurance companies. Internet and digital insurance are becoming significant management tools for insurance policy management. Machine learning and telematics information are used to improve safety monitoring and fraud auditing. In life insurance, unit-linked and equity-indexed contracts have been important innovative products in recent years. Private equity in life insurance markets is also growing, and the effect of interest rates on variable annuities must be better understood.
References

- Publications until 1973


Carter RL (1972) Economics and insurance. PH Press


Greene M (1973) Risk and insurance. South Western


Hammond JD (1968) Essays in the theory of risk and insurance. Scott Foresman


Mahr W (1964) Einführung in die Versicherungswirtschaft. Duncker & Humblot


Moirith RJ (1972) Risk and insurance instruction in American colleges and universities. SS Huebner Foundation for Insurance Education


Pfeffer I (1956) Insurance and Economic Theory. Irwin

Pratt J (1964) Risk aversion in the small and in the large. Econometrica 32:122-136


- Publications after 1973


Eling M (2024) Big data, risk classification, and privacy in insurance markets. Forthcoming in Geneva Risk and Insurance Review


Grace MF (2024) Regulation of insurance markets in the US. In: Dionne G (ed) Handbook of insurance, 3rd edition. Springer


Peter R (2021b) Prevention as a Giffen good. Economics Letters 208:110052
Peter R (2022) When is safety a normal good? Working paper, University of Iowa. Available at SSRN 44147815

Peter R (2024a) The Economics of self-protection: A tribute to EGRIE. Forthcoming in Geneva Risk and Insurance Review


Polemarchakis H (1990) Competitive allocation when the asset market is incomplete. Geneva Papers on Risk and Insurance Theory 15:5-16


Riley JG (1979) Informational equilibrium. Econometrica 47:331-359


Swiss Re (2009) The role of indices in transferring insurance risks to capital markets. Sigma 4


