UNIVERSITÀ COMMERCIALE "L. BOCCONI"
Dottorato di Ricerca in Economia Aziendale e Management
XV Ciclo

RISK-BASED CAPITAL AND MARKET DISCIPLINE IN THE INSURANCE SECTOR

Dottorando: Giacomo Nocera

Abstract

This work aims at showing the peculiar features of market discipline in the insurance market. Market discipline can have different meanings and different kinds of market discipline can be distinguished, according to the markets, the firms, the stakeholders involved. Insurance market discipline works mainly as a (direct) product market discipline, meaning that policyholders' preferences force insurers to comply with their requirements, by limiting their insolvency risk or accepting lower premiums in exchange for a higher risk exposure. Empirical evidence shows that market discipline mechanisms are at work at least in certain lines of business.

En effort to develop a microeconomic model which accounts for the existence of such a market discipline is made. The object of the model is to show that in an insurance market where policyholders are averse towards the default of the insurers and are able to assess such riskiness, insurance demand would also depend on the financial quality of the insurers. Within this framework, the possible impact of market discipline on insurance firms' behaviour is described. If the financial quality of insurers could be observed by policyholders, insurers would also compete on the basis of this perceived quality. They would act in order to conform their riskiness to policyholders' preference, limiting their intrinsic risk or increasing their surplus. In this framework the optimal amount of an insurer's surplus depends on policyholders' preferences regarding insurer's riskiness. Surplus becomes the easiest (but costly) tool that an insurer can use to maximise the probability to meet policyholder's preferences. Therefore, surplus would act as a risk-based capital, since its optimal amount would depend on both the market preferences and the insurer's intrinsic risk.

Contents

Chapter II

Chapter III

Introduction

This work aims at showing the peculiar features of market discipline in the insurance market. This objective stems from the current debate about the solvency regime which regards banking, securities and insurance firms and which in the EU discussion about insurance firms has been named Solvency II.

The subject is of considerable interest and importance. The Basel proposal, with its market discipline pillar, and the question of whether banks should be required to issue subordinated debt to promote market discipline, the on-going formulation of risk-based capital requirements for E.U. insurers and a National Association of Insurance Commissioners (NAIC) proposal to increase significantly the stringency of risk-based capital requirements for U.S. insurers have been widely debated.

The solvency regime, as originally conceived in the Basel discussions on capital requirements for international banks, should consist of 3 pillars: financial strength, the supervisory review process and market discipline.

The first pillar, financial strength, is the quantitative capital requirements. In the current solvency regime these requirements to a large extent do not reflect the risks assumed by the specific firm. The solvency II discussions are aiming at setting the solvency requirement according to all the types of risks that a firm has assumed. It is anticipated that insurers may use their internal models for setting solvency requirements in order to support the development and operation of costly risk management systems. However the models must fulfil certain criteria in order to be accepted by the supervisor. Smaller firms that do not want to build their own models may use more simple models, which have generally been approved by the supervisor.

The second pillar, supervisory review process, consists of the supplementary instruments in order for the supervisor to assess and potentially reduce a firm's risk profile.

The third pillar is market discipline. The aim of this pillar is to strengthen market discipline by creating transparency. Requirements to disclose *inter alia* the financial statements of the insurance firms are a part of this, but it is equally important that the financial reports of the insurance firms are readily comparable.

If both practitioners and theoreticians are making a constant effort to assess which the features of the first two pillars should be, up till now little has been said about the actual part that the market discipline could perform in within the framework of the forthcoming solvency regime.

A wider debate and a consequently broader literature about market discipline has grown in the banking industry, but insurance is a so different industry from banking, that any effort to apply the same insights to the insurance industry start would be misleading. A more consistent approach should start by recognising first what makes insurance so peculiar and to which extent a general concept as market discipline can apply to the insurance market in the framework of regulation and supervision.

The work is organised as follows. Chapter I provides a definition of market discipline, starting from a general concept, suitable for every kind of firms, and ending by narrowing this concept in order to take into account the specific features of financial intermediaries and, among them, of insurance firms.

Market discipline can have different meanings and different kinds of market discipline can be distinguished, according to the markets, the firms, the stakeholders involved. In Chapter II the main features of the market discipline mechanism in the insurance market are outlined. Insurance market discipline works mainly as a (direct) product market discipline, meaning that policyholders' preferences (namely: policyholders' risk aversion) force insurers to comply with their requirements, by limiting their insolvency risk or accepting lower premiums in exchange for a higher risk exposure.

In a sense, market discipline is more effective where policyholders prove to be more sensitive towards the insurers' risk of default, the costs borne by the insurers in case of any increase of their risk of default are high and insurers are adequately of transparent. Empirical evidence shows that market discipline mechanisms are at work at least in certain lines of business.

If some market discipline mechanism exists in the insurance sector, the direct research question is to find out its rationale, trying to point out its main requirements, its implication for the insurers' behaviour and its possible impacts on the definition of an efficient supervisory system.

En effort to develop a microeconomic model which accounts for the existence of such a market discipline is made in Chapter III. The object of the model is to show that in an insurance market where policyholders are averse towards the default of the insurers and are able to assess such riskiness, insurance demand would also depend on the financial quality of the insurers. In such a framework insurance discipline would exert two correlated effect: a price effect (*i.e.*, the higher the risk of default of the insurer, the lower the insurance premium policyholders are willing to pay) and a quantity effect (*i.e.*, the higher the risk of default of the insurer, the lower the insurance demanded).

Chapter IV follows the framework defined in Chapter III in order to describe the possible impact of market discipline on insurance firms' behaviour. Given the policyholders' preferences, the possible reactions of insurers are described both in case of a single insurer negotiating with a representative policyholder and in the case of a number of insurers competing to maximise their utility. If the financial quality of insurers could be observed by policyholders, insurers would also compete on the basis of this perceived quality. They would act in order to conform their riskiness to policyholders' preference, limiting their intrinsic risk or increasing their surplus. In this framework the optimal amount of an insurer's surplus depends on policyholders' preferences regarding insurer's riskiness.

Surplus becomes the easiest (but costly) tool that an insurer can use to maximise the probability to meet policyholder's preferences. Therefore, surplus acts as a risk-based capital, since its optimal amount depends on both the market preferences and the insurer's intrinsic risk. It could be said that it is the market discipline mechanism (as described in Chapter IV) that correlates surplus to insurer's intrinsic risk.

According to Harrington (2004*b*), where a market discipline mechanism works effectively,

> *the benefits of stringent risk-based capital requirements, which will have limited accuracy under the best of circumstances, will be small. The costs from distorting decisions of sound insurers will be comparably large. As a result, risk-based capital requirements should be designed not to bind the decisions of most insurers, and simple requirements are likely to be as effective as more complex rules.*

Obviously, in order that this result held, the following three prerequisites should be met:

(1) policyholders must be sensitive to insurers' insolvency risk;

(2) market responses to changes in the insurer's risk profile need to have cost implications for the firm and its managers;

(3) policyholders are able to assess insurers' insolvency risk.

Any degree of opacity in the opaqueness of insurance firms' financial data relaxes the effectiveness of market discipline and reduces the significance of a statement as the one quoted above.

CHAPTER I

Market discipline: an introduction

I.1. Introduction

While market discipline is a widely-used term, especially in the context of supervising and regulating financial institutions, it is frequently not precisely defined. The aims of this chapter are: (1) pointing out a broad definition of market discipline, (2) identifying its main features, and (3) recognising the different kinds of market discipline.

Firms' behaviour is naturally affected by a number of different stakeholders. Each of them carries his own claims and requirements, which reflect his own preferences. Prompted by claims and requirements which often diverge (as they stem from stakeholders with preferences that may differ), a firm should act in order to maximise the aggregate utility of their stakeholders, trying to find the most adequate balance between stakeholders' claims and requirements. Certain stakeholders' claims and requirements are sometimes so pressing that any failure to meet them could be prejudicial to the firm's capability to attain its institutional objectives in a properly way.

Such kind of failures leads the involved stakeholders to react and punish the firm. One obvious reaction is to leave the firm or asking some kind of compensation.

For instance, if a firm acted in such a way that its default risk increases, its suppliers or creditors could decide either not to serve it anymore or to increase the cost of supplying or the interest rate. Similarly, if the firm offered low quality goods or services, its clients could decide either to address themselves to another firm or to ask for a lower price.

Summing up, each stakeholder exerts some influence on the firm, *i.e.*, the firm is exposed to a sort of stakeholders' discipline. When this discipline can be led back to a market, it can be named as *market discipline*.

I.2. A broad definition

Broadly speaking, market discipline means that markets provide signals that exert some kind of influence over firms' behaviour. It is a general definition, since it does not specify the kind of markets, of signals, or influence exerted.

In fact, since a number of stakeholders exists, different kinds of market discipline could be identified: the product market discipline, the factor market discipline, the financial market discipline, etc.

Regarding signals, usually market discipline mechanism makes use of the terms of trade (*e.g.*, prices, interest rates, salaries, etc.) and quantities (*e.g.*, sales, volume of borrowing, jobs, etc.) demanded as the main *signals* that can discipline firms' activity[1].

These signals can have a twofold effect: internal (or direct) and external (or indirect). The internal effect refers to the impact that market discipline signals have on the behaviour of the firms themselves. Terms of trade and quantities affect directly firms' revenues and cash flows, so they can force firms to act in such a way that these signals are as favourable as possible.

The external effect refers to the chance that market discipline signals are used by third parties (*e.g.*, supervisors) as a source of information regarding the firms' inclination to meet certain stakeholders' preferences.

[1] Obviously, prices and quantities are high closely related signals.

Those two effects correspond, respectively, with the following two different kinds of market discipline highlighted by Flannery: *direct and indirect market discipline*[2].

Direct market discipline can be defined as the process whereby the market discipline signals affect the economic and financial position of a firm.

Indirect market discipline can be defined as the process whereby the market discipline signals are used as a means for third parties (usually: supervisors) to improve their monitoring and controlling tasks. The existence of an indirect market discipline can arise from the inadequacy of the direct market discipline. Where direct market discipline does not work properly other kinds of protection are needed to safeguard certain stakeholders' interest. An indirect market discipline can be a way to strengthen stakeholders' protection, especially when stakeholders can be assumed to be "weak", as creditors and customers.

I.3. Different features of market discipline: product vs financial market discipline

The previous broad definition of market discipline is in force in each kind of market, provided that it works efficiently. However, market discipline can exhibit specific features according to the specific nature of the market. A remarkable difference exists between product markets and financial markets, therefore between *product* and *financial market discipline*.

I.3.1. Product market discipline

Usually a product market is assumed to work efficiently when it is a competitive market, so the previous definition of market discipline could be regarded as a *competition-driven market discipline*, meaning that competitive markets does not leave much room to those firms that fail to act according to their rules.

[2] See Flannery (2001).

The most important rule the firms have to comply with is that markets exert a preference-driven demand.

The meaning is quite clear: a firm will not find any room in the market if it fails to meet market's (*i.e.*, consumers') preferences. Such a failure is usually caused by a lower quality of the firm's products, compared with its competitors'[3].

If the firm failed to meet the market's preference a competitive market would address itself to other firms, reducing the demand of the firm's products and decreasing, in turn, the price of the firm's products till the new lower prices reflect adequately the lower quality of the firm's products.

I.3.2. Financial market discipline

One of the main differences between product markets and financial markets lies in the way the market evaluates the firms' inclination to meet the consumers' preferences. In product markets the evaluation of the (quality of the) product *per se* is usually sufficient to appraise firms' inclination to meet consumers' preferences.

Problems arise when the services of the two counterparts do not occur at the same time, *i.e.*, when the firm-consumer relationship is a long term one. Such a situation typically occurs when the consumer is a firm's creditor. In this case the valuation of the firm's goods or services is not enough, but a valuation of the firm's trustworthiness is also needed.

Therefore, in financial markets an evaluation of the (quality of the) product (*i.e.*, a financial instrument) requires an evaluation of the (quality of the) issuer firm itself[4]. The quality of a financial instrument is closely linked to the firm's earning power and creditworthiness. It could be said that the quality of a financial instrument depends not only on the obligations written in the contract but also –

[3] Here the word "quality" (referred to a product) means the way a product meets certain needs. A high quality product is able to meet adequately those needs. Hence, as referred to certain needs, quality has an objective nature. However, since needs and preferences are subjective and mutable, quality has also a subjective and mutable nature.

and especially – on the reliability of those obligations. It requires that the market is able to evaluate the firm's earning power and creditworthiness, which requires, in turns, an adequate degree of firms' transparency[5].

As a consequence, competition is not a sufficient condition for a financial market to exert a market discipline. An adequate degree of information transparency is also needed. In a sense, financial market discipline could be regarded mainly as *information-driven*. It is clear that in such a context the market discipline becomes more complex, since the quality evaluation becomes a difficult matter.

I.4. Market discipline for financial intermediaries

Usually a firm operates in both product and financial markets. Therefore, it undergoes different kinds of market discipline mechanisms[6].

A singular case is when the firm is a financial intermediary. As pointed out by Merton:

> *Intermediaries are different from other business firms because they create explicit liabilities whenever they sell their products*[7].

As far as market discipline is concerned, a financial intermediary differs from others firms, because the market discipline mechanism always requires the evaluation of the firm's quality.

In other words, if referred to financial firms, market discipline always means that financial markets exert such an influence that force firms to conform their quality to market's preferences. Consequently, a further specification of a financial firm's quality within the framework of market discipline is required.

[4] Sometimes an evaluation of the firms' quality could be also required in product markets. As an example, consider the case of a sale relationship. The buyer could be interested in the firm's quality if the firm provides also spare parts or after-sales customer service.

[5] In fact the difficulty of the evaluation of this reliability (arising from both information shortage and asymmetry) is one of the causes of the existence of financial intermediaries and supervision.

[6] Cf. Llewellyn, Mayes (2003), p. 9.

[7] See Merton (1989), p. 242.

Usually *quality* refers to the financial firm's solvency.

In his seminal work on market discipline, for instance, Lane asserts that

> *market discipline means that financial markets provide signals that lead borrowers to behave in a manner consistent with their* solvency[8].

Lane's definition of market discipline is much more detailed then the broad one suggested above, since it refers to a specific market (*i.e.*, the financial market) and a specific kind of influence exerted (*i.e.*, the achievement of a consistent degree of risk of default). However Lane's definition is not restricted only to financial intermediaries, but it refers to any borrower (*e.g.*, national governments).

According to Lane, the main market discipline mechanism descends from the relationship between interest rate and volume of borrowing: as borrowing increases, the market initially insists on a higher interest rate spread and eventually excludes the borrower from further lending altogether[9]. Interest rate would act as the signal which reveals a likely increase of the borrower's riskiness.

Lane identifies four general conditions for market discipline to be effective.

(1) Financial markets should be free and open (so that interest rates can respond to the level and nature of borrowing); (2) adequate information must be available about the borrower's existing debts and the prospects of repayment; (3) there should be no possibility that lenders will be bailed out in the case of an impending default; (4) the borrower should respond to the market signals before being excluded from the markets[10].

In fact, although the term market discipline can be referred to a generic market, it is usually used within the limited framework of financial markets. Therefore, the term market discipline usually refers to peculiar markets, the financial markets, with the peculiar mechanisms and signals described above. In addition, the

---

[8] See Lane (1993), p. 55.

[9] Lane identifies two other market discipline mechanisms, which can apply when the borrower is a national government: (1) the relationship between interest rate and exchange rates (higher interest rates may be associated with anticipated exchange rate depreciation) and (2) capital flight, which often occurs in response to policies perceived to be unsustainable, policies that result in inflation, exchange rate depreciation, and perhaps heavy taxation and causes domestic borrower to lose access to their country's savings. Conversely, when sustainable policies are adopted, flight capital frequently returns home. See Lane (1993), pp. 57-60.

[10] *Ibid.*, p. 61.

suggestion to make use of the market discipline mechanism[11] to strengthen the efficiency of banking supervision has restricted the common meaning of market discipline so that it has ended up by applying almost exclusively to financial firms (and, particularly, to banks[12])

However, even if this kind of market discipline seems to work only in one way, further considerations can be made. In fact, if one looks more carefully at the financial firm's counterparts, two different kinds of market discipline for financial firms can be identified.

Following Merton (1989) two kinds of financial intermediaries' counterparts exist: customers and investors.

A distinction between customers and investors of nonfinancial firms is rarely necessary, because it is generally obvious.

Few would confuse the customer who buys a car from an automobile firm with the shareholder, lender, or other investor who buys its securities[13].

An analogous distinction is not always so obvious in the case of financial intermediaries. In fact if

no one would confuse a customer who changes money at a bank or takes out a loan from it with an investor who owns shares in the bank (...) the customers of many types of intermediaries receive a promise of services in the future in return for payments to the firms now. Financial services of this type such as insurance and retirement annuities usually involve payments to the customer of specified amounts of money, contingent on events and the passage of time. Those promised future services are liabilities of the firm, both economically and in the accounting sense. Since investors in the firm also hold its liabilities, the distinctions between customers and investors is not always so clear for such intermediaries[14].

[11] Both direct and indirect.

[12] See, for instance, Board of Governors of the Federal Reserve System (1999), Bliss (2001), Bliss, Flannery (2001), Flannery (2001), Sironi (2001).

[13] Cf. Merton (1997), p. 2.

[14] *Ibid.*, pp. 2-3.

The major liabilityholders of those financial firms that act as a principal in the ordinary course of business are their customers[15]. In fact

although intermediaries do act as agents in some transactions, their primary function is to act as principals and provide financial instruments and products that cannot be efficiently supported by trading in organized secondary markets. The purchasers of these products are therefore de facto *liabilityholders of the intermediary. Indeed, as we all know, the vast bulk of a typical intermediary's liabilities are held by customers*[16].

The difference between customers and investors is very subtle (if it is referred to a financial intermediary) and it can be found in the different individual attitude towards the intermediary's quality. According to Merton, customers who hold the intermediary's liabilities

are identified by their strict preference to have the payoffs on their contracts as insensitive as possible to the fortunes of the intermediary itself[17].

The efficiency of customer contracts is diminished if they are exposed to default[18].

In a sense, from the point of view of the customers, the quality of a financial intermediary is closely linked to its *default risk* or, more precisely, the prospect of its *contract default*. In fact, as pointed out by Merton

creditworthiness or default risk is a critical issue for all business firms – and for households as well. However, for financial intermediaries with principal business that involve issuing contingent-payment contracts to their customers, creditworthiness is the central *financial issue. The prospect of a future default by an intermediary on contracts to its customers can significantly reduce* ex ante *efficiency*

[15] See Merton, Perold (1993), p. 16. Principal activities of financial intermediaries can be assed-related (as in the case of lending), liability-related (as in the case of deposit-taking and writing guarantees, *e.g.*, insurance) or both (as in the writing of derivatives for customers).
[16] See Merton (1989), p. 242.
[17] See Merton (1997), p. 3.
[18] *Ibid.*, p. 4.

to those contracts and thereby substantially reduce the effectiveness of the main economic function served by the intermediary[19].

A key requirement for the success of any financial intermediary is its ability to control both the actual and perceived default risk of its customer-held liabilities[20].

Going back to the definition of the market discipline mechanism for financial intermediaries, it can be said that where customer relationships prevail market discipline works in such a way that financial firms are forced to limit their default risk, since, as pointed out by Merton,

because of the substantial size and long duration of many financial contracts such as annuities and life insurance, default risk is a first-order issue for customers of financial intermediaries. Thus, the success of a financial intermediary depends not only on charging adequate prices to cover its production costs, but also on providing adequate assurances to its customers that promised payments will be made[21].

Investors are less credit-sensitive than customers, provided they are compensated appropriately for the risk they incur, since, as pointed out by Merton

investors in the liabilities issued by an intermediary (e.g., stocks or bonds) expect their returns to be affected by its profits and losses[22].

Also any investor exhibits an aversion towards the *contract default* and this aversion gives rise to a market discipline mechanism.

However, if it can be assumed that both customers and investors exhibit an aversion towards the intermediary's risk of default, they will probably exhibit different extents of such aversion. The difference would concern the extent of the price/quality trade-off. If investors could be willing to invest in low quality intermediaries in exchange for an adequately high return, customers could not be willing to buy liabilities which exceed a certain risk.

---

[19] Cf. Merton (1995), p. 24.
[20] *Ibid.*, p. 36.
[21] Cf. Merton (1997), p. 4.

Consequently the two market discipline mechanisms (the *investor-driven* and the *customer-driven*) will produce the same kinds of signals (*i.e.*, price – or returns – and quantities) but in different ways. In particular, price signals should be stronger where *investor-driven* market discipline mechanisms prevail, while quantity signals should be stronger where *customer-driven* market discipline mechanisms prevail.

Consequently a distinction between a *product market discipline* and a *financial market discipline* holds also with reference to financial intermediaries, but they both stem from the intermediary's quality[23] and they differ only because of the kinds of counterparts involved (respectively, customers and investors) and the features of the discipline.

I.5. Market discipline for insurance firms

Having defined the main features and kinds of market discipline, the following research questions could arise:

(1) is there any market discipline mechanism in the insurance sector?
(2) which features does market discipline exhibit in the insurance sector?
(3) which are the main implications of market discipline for the design and implementation of solvency regulation in the insurance sector?

Those questions are answered in the following chapters.

Particularly, Chapter II describes the peculiar nature of market discipline in the insurance sector. According to the previous distinction, it can be said that market

[22] Cf. Merton (1997), p. 3.

[23] In fact, market discipline always requires that the market prices of the financial instruments issued by the financial firm (held either by investors or customers) contain accurate and timely information about the firm's risk.
Particularly, as far as principal financial firms are concerned, market discipline always requires that the market prices of the financial instruments issued by the financial firm (held either by investors or customers) contain accurate and timely information about the insurer firm's risk.

discipline in the insurance sector is basically a *product market discipline*. The main prerequisites for the effectiveness of such a market discipline are presented and, finally, some empirical evidence of market discipline in the insurance sector is reviewed.

Chapter III explains the microeconomic foundations of the market discipline mechanism. The assumption that policyholders are averse towards the default risk of the insurers in the framework of assessment of policyholders' propensity to insure leads to the formulation of an insurance demand affected, negatively, by the insurers' riskiness.

The main implications of such a market discipline for solvency regulation are presented in Chapter IV. Since market discipline forces insurers to conform to policyholders preferences, its mechanisms should be strengthened and perfected. A strengthening is justified by the fact that market discipline may turn to be more effective than any other solvency tool, while a perfecting could be required because an imperfect market discipline mechanism could lead to biased and sub-optimal results.

CHAPTER II

Market discipline in the insurance sector

II.1. Introduction

As shown in Chapter I, while a co-existence of investors and customers (and, therefore, a co-existence of financial and product market discipline) occurs in any firm, it exhibits special features if it refers to financial firms.

Since an insurance firm can be assumed as a financial firm acting as principal, the co-existence of two kinds of market discipline (*investor-* and *customer-driven*, or product and financial), can be postulated also in the case of insurance firms. In fact, such a specification is consistent with the definition of (direct) market discipline in the insurance sector raised by Harrington. According to his definition, (direct) market discipline means the extent to which product and capital markets discipline risk-taking by insurers[24].

Financial theory views the insurance firm as a levered corporation with equity capital and a special kind of debt, raised by the insurer by issuing insurance contracts. Consequently, insurance firms' investors and customers correspond, respectively, to shareholders and policyholders.

---

[24] Harrington (2004*b*).

In spite of the co-existence of an *investor-driven* and a *customer-driven* market discipline, insurance sector market discipline refers mainly to the *product market discipline*, *i.e.*, to the customer-driven (that is: policyholder-driven) market discipline.

Harrington defines product market discipline in the insurance sector as

> *the extent to which demand by policyholders is sensitive to insolvency risk and thereby motivates insurers to manage their risk*[25].

On the contrary, *financial market discipline*, *i.e.*, the investor-driven (that is: shareholder-driven) market discipline has been practically disregarded. The rationale can be found in the peculiar nature of the insurance investor-driven market discipline, which *de facto* is a shareholder market discipline. In fact, as pointed out by Kwan[26]:

> *the limitation of bank stock prices is that they are not straightforward to interpret because movements in bank stock prices can be driven not only by changes in bank asset value but also by changes in bank asset risk. To see the latter, because stockholders have claims on all of the firm's cash flow after paying off bondholders, increasing a bank's asset risk would benefit the stockholders at the expenses of the bondholders, since the stockholders would get all the upside risk, while the bondholders would bear only the downside risk. Moreover, stockholders' incentive to take excessive risk grows as the bank's capital situation worsens, which is especially problematic for bank supervision: at the very moment when the surveillance of weak banking institutions becomes crucial, the stock market signal may be most susceptible to conveying conflicting information*[27].

Therefore, for obvious reasons any effort to strengthen the market discipline in the banking sector has been focused on the market discipline mechanism arising from the pricing behaviour of banks' uninsured debt, namely the subordinated debt[28].

---

[25] Cf. Harrington (2004b).

[26] Kwan refers his analysis to banks, but his remarks hold also with regard to insurance firms.

[27] See Kwan (2002), p. 2.

[28] Much less attention has been paid to bank deposit (*i.e.*, the production side of the bank). The reason can be found in the peculiar nature of those instrument that do not make them right to the

II.2. Prerequisites for an effective market discipline in the insurance sector

If market discipline in the insurance sector refers mainly to the extent to which demand by policyholders is sensitive to insolvency risk, and thereby motivates insurers to manage their risk, the three following main prerequisites for an effective market discipline should hold:

(1) policyholders must be sensitive to the insurers' risk of default;
(2) market responses to changes in the insurer's risk profile need to have cost implications for the insurance firm and its managers;
(3) policyholders are able to assess insurers' insolvency risk.

In a sense, the higher policyholders' sensitivity to the insurers' risk of default, the costs borne by the insurers in case of any increase of their risk of default or the degree of insurers' transparency, the stronger the market discipline.
The rest of the chapter aims to show to which extent those conditions are met, using both a review of the theoretical arguments and the empirical evidence.

II.3. Theoretical evidence

II.3.1. Policyholders' sensitivity to the risk of default of the insurers

The idea that policyholders are averse towards the insurers' risk of default and therefore are concerned not only with the premium charged but also with the insurers' financial quality is not new.

---

market discipline purpose. In fact, as showed by Blum (2002) and Cordella and Yeyati (1998), if bank deposits are uninsured and the bank's risk choice is observable by depositors, the bank's risk choice will be efficient. This is so because banks take account of the impact of their risk choice on depositors since these will demand higher compensation should the bank incur higher risk.

In fact, policyholders exhibit all the requisites that distinguish typical financial intermediaries' customers, a class of individuals who, according to Merton and Perold,

> *prefer to have the payoffs on their contracts as unaffected as possible by the fortunes of the issuing firm. Hence, they strongly prefer firms of high credit quality (...). This means that A-rated firms, for example, can generally raise the funds they need to operate, but at a disadvantage in competing with AAA-rated firms in business such as underwriting insurance or swaps*[29].

The customer nature of policyholders is explicitly pointed out in the following example suggested by Merton.

> *The function served by a life insurance policy is to provide its beneficiaries with a specified cash payment in the event of the insured party's death. That function is less efficiently performed if the contract calls instead for the death benefit to be paid in the joint event that the insured party dies and the insurance company is solvent. Even if the insurance company offers an actuarially fair reduction in the price of the insurance to reflect the risk of insolvency, a risk-averse customer would prefer the policy with the least default risk. Indeed, on introspection, I doubt that many real-world customers would consciously agree to accept non-trivial default risk on a \$200.000 life insurance policy in return for a large reduction in the annual premium, say from \$400 to \$300. Such results obtain* even *in theoretical models in which the customer has all of the relevant information necessary to assess the default risk of the insurer. In most real-world cases, the customers do not have the relevant information, and this fact* a fortiori *makes the potential welfare loss from customer-contract default even greater*[30].

Customers of financial firms are therefore concerned with the expected losses on their contracts with the firms, *i.e.*, with the so-called contractual performance risk. The estimated losses depend on the probability of default of the firm, the loss rate

---

However, if deposits are insured or the bank's risk choice is not observable by depositors, the bank will increase risk at the expense of depositors.

[29] See Merton e Perold (1993), p. 16.

[30] See Merton (1997), p. 3.

given default and the exposure at default. Hence, naturally, the policyholders, as customers of financial firms, are interested in purchasing insurance cover from insurers with a low probability of default and a low loss rate given default – *i.e.* with a prudent risk profile.

If insurers' customers – especially retail policyholders – face contractual performance risks which are large relative to their wealth, it is obvious that, once premiums are paid, policyholders worry whether their future policy claims will be honoured swiftly and fully[31].

Froot suggests some reasons to believe that customers act as though they were more risk averse than investors, identifying two main kinds of explanations behind the greater product market sensitivity of customers, one is behavioural and the other is rational.

The *behavioural* explanation refers closely to the "certainty effect" described in Kahneman and Tversky's Prospect Theory[32], according to which individuals prefer certainty, and will pay more to reduce uncertainty by one unit the lower is the probability of a bad outcome. In other words, when an outcome is certain and it becomes less probable, this has a greater impact than when the outcome was merely probable before the probability was reduced by the same amount. Thus a reduction of probability from 100% to 90% has an heavier psychological impact than reduction of probability from 50% to 40%, even if the absolute reduction is the same (10%). What makes this argument especially specific to insurance – since the same will hold true of an individual's desire to eliminate the last portion of risk in a portfolio of securities – is the fact that while, for multiple reasons, portfolios often are managed at relatively high levels of risk, suggesting that the price of risk reduction is low, insurance claims, by contrast, have relatively low risk of performance failure, and therefore, the price of risk reduction should be higher.

The *rational* explanation comprises several arguments. The first argument refers to the higher costs which customers incur to manage the default risk of financial intermediaries compared to the lower cost incurred by investors. As pointed out

[31] As pointed out by Froot (2003), p. 2, commercial bank customers are less exposed to the default risk of their banks, because they are protected by deposit guarantees as depositors and are not exposed to bank performance risk as borrowers.

by Merton[33], it is true that as investors can eliminate or at least reduce significantly the overall impact of the default risk of the firms they have invested in through a wise diversification, customers could be able to reduce the effect of this default risk either by trading in the securities of the firm (e.g., an insurance company) or by entering into a large number of tiny contracts (e.g., an insurance policies) with many different companies. But it is also true that these strategies of "hedging" and "diversification" could be possible only if markets were complete and frictionless. «*The very economic role of the intermediary* – Merton claims – *is to service those entities (its customers) who cannot trade efficiently and who cannot enter contracts costlessly*». Therefore, higher frictional costs of diversification[34] translate into effective risk aversion that is higher for customers than investors. A second argument is that customer losses occur predominantly when customers' wealth is low and, therefore, the marginal value of wealth is high. If absolute risk aversion is declining in wealth (as many suspect it is), then a customer will be more averse to an insurer's failure to perform than to a debtor's failure to perform, even if the performance failure is of equivalent size. A third argument is what Froot calls the "hassle" factor. Customers face bores and troubles in dealing with a failed or failing insurer. Policies must be cancelled, claims and premium rebates must be pursued more vigilantly or even sacrificed, and a new coverage must be found. As the insurer's risk increases, the expected cost of these hassles is greater for customers than investors[35].

The combined effect of these arguments and the resulting large contractual performance risk borne by insurance customers should make demand sensitive to the insurer's risk, where this sensitivity of customer demand could appear as either a change in the price an insurer can charge or a change in the quantity that they insurer can sell.

---

[32] Cf. Kahneman, Tversky (1979).
[33] See Merton (1995), p. 34.
[34] Froot (2003), p. 4, claims that, in the case of insurance, the costs of diversification would arise from the fact that a considerable information transfer between buyer and seller is needed, since insurance contracts are complex instruments for buyers to understand and value. And, in the event of a claim, it is costly for the buyer to notify multiple insurers, coordinate their monitoring requirements, and negotiate settlements.
[35] *Ibid.*, p. 5.

The policyholders' aversion to the risk of default of the insurer is explicitly postulated in the works of Cummins and Sommer (1996), Cummins and Danzon (1997), Zanjani (2002*a*) and Cummins and Mahul (2003).

Cummins and Sommer assume that policyholders are averse to insolvency risk, *i.e.*, «*they do not view insurance solely as a financial asset but purchase insurance in order to mitigate the risk of loss*»[36]. They formalise their assumption about insurance demand by introducing an insurance demand function which is inversely related to both the unit price of insurance and the standardised insolvency put value[37], and therefore it is not infinitely elastic. The fundamental premise is, obviously, that policyholders have full information about insurer insolvency risk and adjust their demand for insurance accordingly.

Similar assumptions can be found in the work of Cummins and Danzon, who ground their model on the idea that, since insurers are subject to default risk, insurer liabilities are analogous to risky debt and can be modeled using the standard risky debt model of the firm. Therefore, policyholders' demand for insurance depends on the expectations about the insurer's financial quality, as measured by the insurer's insolvency put option per dollar of liabilities[38]: the demand for insurance is positively related to financial quality and imperfectly price elastic in the short run. As in the work of Cummins and Sommer, the expectations about the insurer's financial quality are based on observable leverage ratios and possibly other characteristics.

Moreover, Zanjani also assumes that consumers care about the financial condition of insurers. Even in those countries where guaranty funds are in place, policyholders still bear some of the burden of insolvency, because recoveries are usually capped, delayed, and subjected to additional deductibles. In addition, guaranty funds do not cover some classes of policyholders (notably, insurance companies). Zanjani assumes that policyholder's utility in a given market segment i, $U_i.(p_i, q)$, depends (negatively) on the price of the insurance coverage in that segment, p_i, and (positively) on the quality, q, of the product, where "quality"

[36] See Cummin, Sommer (1996), p. 1074.
[37] This standardised value is equal to the premium of the insolvency put is divided by the total amount of liabilities.
[38] See Cummin, Danzon (1997), p. 13.

refers, again, to the financial soundness of the insurance company issuing the policy.

Finally, Cummins and Mahul analyse the design of an optimal insurance policy under the assumption that insurers have non-zero probability of default and the buyer is aware that the probability of insurer's default is non-zero.

II.3.2. Implications on the insurance firms' costs of changes in risk profile

It can be assumed that effects of any increase of the insurer's default risk are twofold. One effect comes from the policyholders' risk sensitivity described above. Epermanis and Harrington describe this effect as follows:

> *when customer demand is risk sensitive, financial institutions with higher insolvency risk will receive less favourable terms of trade, such as having to promise higher yields on bank deposits, CDs, or subordinated debt, or having to accept lower prices on insurance contracts. Deterioration in an institution's financial condition and an attendant increase in its insolvency risk likewise should produce a loss of customers and adversely effect the terms of trade with customers that remain*[39].

Indeed, the most relevant market discipline mechanism at work in the insurance sector lies in this effect. However, any change in the firm's risk profile has another effect, which do not operate through policyholders' behaviour, since insurers, like any other firm, are naturally concerned about the expected costs of financial distress.

Harrington (2004*b*) identifies three main items that can affect insurers' behaviour and motivating a direct risk reduction action: (1) the potential loss of franchise

[39] Epermanis and Harrington (2004).

value, (2) the effects of financial distress on other stakeholders, and (3) the costs arising from bankruptcy and control changes.

The potential loss of franchise value

Insurance production and distribution often involve the creation of firm-specific assets (franchise value) that could diminish or evaporate if the insurer experiences financial distress. As emphasised in both the banking and insurance literatures, franchise value can provide incentives for adequate capitalisation and other forms of risk management even when the risk sensitivity of demand is dulled by guarantees or information problems. Risk sensitivity of demand amplifies those incentives because financial difficulty will then reduce both new and renewal business.

Insurers' firm-specific assets arise in four main ways. First, attracting and providing coverage to new customers typically requires relatively high up-front costs, which insurers expect to recover from higher margins on renewal business. Thus, renewal premiums often include quasi-rents as a return for the initial investment in creating the customer relationship. Second, insurers may obtain valuable private information over time about their continuing policyholders' risks of loss. Third, some insurers make substantial investments in developing a brand name or reputation for quality service. Fourth, some insurers' investments in underwriting expertise and distribution systems generate valuable growth opportunities. In each case, much of the value of these investments could be lost in the event of financial distress or insolvency. Loss of quasi-rents and/or growth opportunities is especially likely in insurance given that unexpected reductions in assets values or increases in claim costs can affect many insurers at once, thus reducing the ability of other insurers to acquire the operations of distressed companies.

The effects of financial distress on other stakeholders

Policyholders and shareholders do not exhaust the set of agents who have something at stake in the relative success or failure of the firm. According to Llewellyn (2002), those who are interested in the process of observing the behaviour of the firm, forming judgements and behaving in the light of it can be described as "stakeholder monitors". They include also board of directors, managers, borrowers, employees, intermediaries, etc.

If there is a chance that their "firm-specific" investment could be lost because of financial distress, they are likely to require added compensation for the greater risk. Employees could demand higher wages (or reduce their loyalty or they work effort) at a company where the probability of layoff is greater. Managers with alternative opportunities will demand higher salaries to run firms where the risk of insolvency and financial distress are significant.

Insurance intermediaries (exclusive agents or independent agents and brokers) are significantly concerned about the insurer' s solvency.

Exclusive agents primarily or exclusively represent a single insurance company or group of affiliated insurers. They generally make substantial investments in developing a book of business that would be lost if the insurer(s) issuing the policies became insolvent. This implies that exclusive agency insurers (direct writers) will have to provide credible assurance of continued financial viability to attract and develop exclusive agents. Independent agents and brokers arrange or distribute insurance for numerous insurers. Other things equal, they make smaller investments in developing relationships with any one insurer than exclusive agents. However, independent agents and brokers make substantial investments in relationships with their policyholder clients. The insolvency of an insurer could damage those relationships if it imposes costs on policyholders. Insolvencies also impose direct costs on the agent/broker, such as the cost of transferring business to other insurers and can produce costly litigation for alleged failure of the agent or broker to protect policyholders' interests.

Finally, since many insurers issue conventional debt, almost always at the holding company level, an additional category of stakeholder monitors can exist: bondholders. Bondholders are risk sensitive and should press for efficient risk management. In such a case, the more conventional financial market discipline mechanism would work. It creates a set of hard claims that could dissuade excessive risk taking ex ante and limit risk taking ex post, especially compared with opaque, long-tailed claim liabilities that can be further extended by end-game claim settlement strategies. Even though many or most of the issues are effectively subordinated to policyholder claims, the bulk of insurer debt is investment grade (Standard & Poor's rating of Baa or above[40]). Credit ratings on insurance entity debt issues also provide another source of information that can be used by interested parties in assessing insurer financial strength. Hence, although the details may differ significantly from the design features of mandatory subordinated debt proposals for banks, insurer debt finance should provide some of the benefits of market discipline that motivate such proposals.

According to conventional theory, the possible direct bankruptcy costs associated with higher payments to these stakeholders will be borne by issuers *ex ante* and help motivate managers to reduce the probability of bankruptcy.

Finally, an additional kind of monitoring effect can be identified: the monitoring on insurers' riskiness carried on by other insurers. This "peer monitoring" can rise out of either the potential contagion effects of financial problems among insurers[41] and the insurers' incentive to monitor each other and alert regulators to potentially excessive risk-taking by peers due to the post-insolvency assessment funding mechanism of guaranty funds. Lee, Mayers and Smith (1997) call this the *monitoring hypothesis*. This monitoring may preclude insurers from taking on high levels of risk and, hence, from increasing the value of the insolvency option.

The costs arising from bankruptcy and control changes

Bankruptcy causes a firm to incur both direct and indirect bankruptcy costs. Direct bank costs include costs of administration and reorganisation, such as payments to lawyers and court costs. Firms in troubles face considerable interference from supervisors and bankruptcy courts with their operating decisions. And such interference has the potential to cause significant reductions in the value of the firm.

[40] See Harrington (2004*b*).

II.3.3. Policyholders' ability to assess insurers' insolvency risk

The risk sensitivity of the policyholders is a necessary but not sufficient condition for the existence of a risk sensitive of insurance demand. Another condition must hold: policyholders may find not costly or difficult any attempt of assessing (directly or through intermediaries) the insurers' insolvency risk. To this end, the most important variables to be taken into consideration are the insurers' opacity and the cost of monitoring. If either the insurers' opacity and the cost of monitoring are low, it is most likely that the insurance demand will turn out sensitive to insolvency risk.

Policyholders seeking to obtain information on the financial strength of insurance companies are hindered by the complexity of the financial information and the difficulty of predicting insolvency.

Many insurance companies have opaque characteristics, increasing the likelihood that some policyholders will be attracted to low prices or high promised investment returns without recognizing that they could indicate relatively high insolvency risk. The valuation of non-life insurers' unpaid claim liabilities for long-tailed liability and workers' compensation insurance is problematic, as is the valuation of some life insurers' privately-placed fixed income investments and many commercial real estate investments. Many insurance organizations have complex ownership structures. Most insurers have elaborate reinsurance arrangements, sometimes involving dozens of reinsurers. While the primary function of those arrangements is to diversify risk, they reduce transparency and can sometimes be used to hide financial problems.

Policyholders may perform this monitoring activity themselves or, very often, may rely on the monitoring activity exerted by the insurance intermediaries. Insurance agents and brokers expend considerable effort monitoring the financial strength of the insurers they place business with, since these intermediaries expose themselves to potential legal liability if they negligently place business with an insurer that later becomes insolvent.

---

[41] See Fenn and Cole (1994).

Evidence of consumer-driven monitoring includes the widespread use of rating agencies that evaluate claim-paying ability of insurers.

Published ratings provide an opinion on an insurer's financial strength, including its operating performance and its ability to meet its obligations to its policyholders. Insurance ratings do not provide any guarantee against default. They do, however, provide policyholders and investors with some information regarding the likelihood of insolvency.

Insurer ratings basically convey two related types of information: information about the default or claims-paying risk of an insurer relative to other insurers and information about the risk of a given insurer relative to non-insurers.

II.4. Empirical evidence

II.4.1. Introduction

Empirical work on market discipline in banking has analyzed the relation between bank risk and yields and flows of deposits, CDs, and subordinated debt[42]. Recent evidence indicates that credit spreads on uninsured deposits and subordinated debt are generally positively related to bank risk[43]. Goldberg and Hudgins (2002) find that insolvent thrifts experienced declines in the volume and proportion of uninsured deposits prior to insolvency. Crabbe and Post (1994) find that outstanding commercial paper issued by bank holding companies declined following rating downgrades by Moody's.

Several other studies suggest that franchise value can discipline bank risk-taking even when deposits are insured[44]. On the other hand, Billet, Garfinkel and O'Neal (1998) find that banks downgraded by Moody's issued more insured deposits after being downgraded even though their uninsured deposits fell. They suggest that

[42] For an overview, see, for instance, Flannery (1998).
[43] See, for instance, Flannery and Sorescu (1996) and Sironi (2003).
[44] Herring and VanKudre (1987), Keeley (1990), Demsetz, Saidenberg and Strahan (1997).

banks' ability to substitute insured for uninsured deposits undermines market discipline and increases the need for regulation.

II.4.2. Evidence of product market discipline

Compared with the literature in banking, there have been fewer insurance studies along these lines, in part due to data and measurement issues. Non-life insurers, for example, do not report volume of coverage or premium rates per unit of coverage (and the information would not be very comparable across firms if they did). It is very difficult to make inferences about (*ex ante*) price differences across insurers because conditional expectations of claim costs when policies are sold are not observable.

Like other research on insurance prices, Sommer (1996) and Phillips, Cummins, and Allen (1998) employ measures of property/casualty insurers' premiums in relation to realized claim costs as proxies for the price of coverage. They both seek evidence of an empirical relationship between insolvency risk and property-liability premiums. Consistent with risk sensitive demand and market discipline, they find that measurers of property-casualty insurers' premiums in relation to claim costs are negatively related to default risk, implying that prices decreases as default risk increases.

Cummins and Danzon (1997) also find that price is positively related to financial quality. However, the existence of guaranty funds potentially limits the impact of ratings on firm value, to the extent that it shields some policyholders from the full impact of insolvencies.

Zanjani (2002*b*) provides some evidence that life insurance policy termination rates are greater for insurers with lower A.M. Best ratings, although he finds that terminations are not reliably related to rating changes.

Finally, Epermanis and Harrington (2004) provide evidence of changes in non-life insurer premium growth surrounding changes in Best's ratings. Consistent with risk-sensitive demand and attendant market discipline, they generally found economically and statistically significant premium declines in the year of and the year following rating downgrades. Their results also indicate that rating upgrades for relatively low-rated insurers were accompanied by increased premium growth. Consistent with less complete guaranty coverage for commercial insurance compared with personal insurance, they find that abnormal premium growth for downgraded insurers was negatively related to the proportion of an insurer's premiums in commercial insurance. For downgraded insurers with A– ratings and at least 50% of their premiums in commercial lines prior to being downgraded, premiums declined by an average of 15 to 20% during the downgrade year alone. Abnormal premium growth associated with downgrades for A– and low-rated firms also was inversely related to firm size, suggesting greater market discipline for larger firms.

Their findings suggest the existence of market discipline for rated insurers, especially those that write relatively large amounts of commercial insurance, despite guaranty fund protection, the relative opacity of property/casualty insurers balance sheets, and other factors that dull some policyholders' incentives to seek financially strong insurers.

Incentives of policyholders to trade with financially strong insurers may be affected by the existence of guarantees of insolvent insurers' obligations. Particularly it can be assumed that where a state-mandated guaranty funds system exists[45], policyholders may reduce their sensitivity towards insurers' riskiness. Reduced risk-sensitivity of demand by direct insurance buyers could in turn affect the risk-sensitivity of insurers' demand for reinsurance. Limits on the amounts and types of direct insurance losses covered by guarantees, however, should provide many insurance buyers with substantial incentives to deal with financially sound insurers.

Significant limitations on insurance guarantees are efficient because they encourage market discipline. The conventional systemic risk rationale for deposit insurance is theoretically and empirically less applicable to insurance[46]. While

[45] Certain obligations of insolvent insurers are subject to government guarantees in the United States and other major countries, thus protecting some policyholders from the full consequences of insurer default.

[46] Contrary conventional wisdom typically considers the main role of deposit insurance as the prevention of bank runs. The argument linking deposit insurance and bank runs was first formally

subject to debate, most observers agree that systemic risk is relatively low in insurance markets compared with banking, especially for non-life insurance. Low probability events with large losses can simultaneously damage many non-life insurers, and their impact is spread broadly through product line and geographic diversification and especially through reinsurance, which creates material contractual interdependence among insurers. As noted, large shocks can disrupt non-life insurance markets with adverse effects on real activity. Reductions in asset values may affect many life insurers at once, and some policyholders may seek to withdraw funds following negative shocks, perhaps causing some insurers to unload assets at temporarily depressed prices. But shocks to non-life and life insurers do not threaten the payment system, and there is little or no evidence of "pure" contagion associated with major events in either sector, as opposed to rational, information-based flights to quality (see below).

Lee, Mayers, and Smith (1997), for example, provide evidence that asset risk increased for stock property/casualty insurers following the introduction of guaranty funds[47]. Brewer, Mondschean, and Strahan (1997) provide evidence that life insurer asset risk is greater in states where guaranty fund assessments against surviving insurers are offset against state premium taxes and thus borne by taxpayers, which may reduce financially strong insurers' incentives to press for efficient regulatory monitoring.

presented by Diamond and Dybvig (1983), who hold that runs can be a self-fulfilling equilibrium. This idea has been influential in safety net design and has contributed to the view that financial markets are essentially unstable and prone to crises not necessarily backed by fundamentals. Their paper shows that a certain level of deposit insurance can make the "no run" strategy the dominant one, thus eliminating the equilibria with runs. They acknowledge that lender-of-last-resort functions can have the same effect, although they do not explore it formally.

In addition to this role as the prevention of bank runs, deposit insurance protects small depositors. While this may sound less grandiose, it may, in fact, be the more realistic objective. Dewatripont and Tirole (1994) develop a theory of banking regulation based on what they call the *representation hypothesis*. By this they mean that regulation is necessary to represent a large number of small depositors who may find it costly to monitor a bank individually, in particular if their deposits are small. Regulation and supervision will restore adequate incentives for good corporate governance of a bank in the presence of an atomized principal. Deposit insurance arises in this context to protect small depositors by minimizing their losses in case of bank failure. In reality, most deposit insurance systems seem closer to the second approach than to the first. The first approach calls for protection for those most likely to run – arguably, large depositors. The second is consistent with limits on protection per depositor.

II.4.3. *Evidence of financial market discipline*

A number of studies provide indirect evidence of market discipline from insurance company equityholders and that equityholders anticipate a flight to quality following major events that reduce insurers' assets or increase liabilities. Fenn and Cole (1994) and Brewer and Jackson (2002), for example, find that life insurer stock price declines during the commercial real estate and high-yield bond market slumps of 1989-1991 were concentrated among firms with problem assets.

II.4.4. *Evidence of insurance ratings performances*

Insurance company ratings provided by private rating agencies are vitally important to investors, regulators, consumers, insurers, and insurance agents/brokers. Insurers use ratings in their advertising to assure buyers of the firm's strength. Insurance buyers use them in choosing their insurance companies and/or deciding how much they are willing to pay for insurance from particular firms. Brokers and agents often will not recommend coverage with unrated insurers or insurers with ratings below some threshold of financial strength[48], and many corporate insurance buyers require that all their insurers be highly rated. Strong financial ratings give insurers better access to capital markets. Insurer ratings also have a direct impact on the cost of capital, since the primary source of debt capital to insurers is policy liabilities, and lower rated firms will likely have to sell their policies at lower prices compared to higher rated firms[49]. Ratings also provide a valuable tool for regulators in assessing the financial strength of insurers[50].

The failure of highly rated life insurance companies, such as Executive Life Insurance and Mutual Benefit Life Insurance Companies, generated heavy criticism of the leading rating agencies by the insurance press. These bankruptcies

---

[47] See also Lee and Smith (1999) and Downs and Sommer (1999).

[48] Cf. Moody's (1998).

[49] See Doherty and Tinic (1981) and Berger, Cummins, and Tennyson (1992).

[50] Cf. Schwartz (1994).

of very prominent insurance companies raised several questions about the motivations and methods of the rating agency. Ambrose and Seward (1988) find that the ratings of the A.M. Best Company perform as well as financial ratios in distinguishing between solvent and insolvent insurers; neither one is a perfect predictor of insolvency. Ambrose and Carroll (1994) analyse the effectiveness of Best's recommendations into a life insurer insolvency model. Using the recommendation in the year prior to insolvency as the predictor correctly identifies over 90 percent of insolvent insurers but performs poorly in identifying solvent insurers. These results imply that ratings below A or A+ do not necessarily indicate a high probability of insolvency and that parties who rely on these ratings may find that they provide insufficient advance warning of possible financial distress. The financial variables and Insurance Regulatory Information System (IRIS) ratios outperformed Best's recommendations in distinguish between the two groups in a logit model. However, combining all three types of predictors into one model provided the most accurate classification of solvent and insolvent life insurers.

CHAPTER III

Market discipline and insurance pricing

III.1. Insurance as an imperfect tool of risk transfer

The theory relating to insurance pricing often regards insurance contracts as transactions where policyholders exchange uncertain "prospects" with certain ones at the cost of a premium paid to the insurer[51]. Consequently, the theory of the demand for insurance has been based on expected utility theory and an assumed preference for certain losses over uncertain ones of the same expected magnitude[52]. At a more general level, but still part of the demand-for-insurance-as-demand-for-certainty theory, other studies have postulated that the demand for insurance is by "risk averse" policyholders who use insurance to eliminate the risk of loss[53]. According to the body of this theory, the amount of the insurance premium is determined after an economic decision principle has been adopted by each the insurer and policyholder. An opportunity exists for a mutually advantageous insurance policy when the premium for the policy set by the insurer is less than the maximum amount that the prospective policyholder is willing to

[51] See, for instance, Borch (1974), Bühlmann (1970), Beard *et al.* (1984) for further discussion.
[52] E.g., Friedman and Savage (1948), Arrow (1963).
[53] E.g., Arrow (1963), Pratt (1964), Mossin (1968), Borch (1974), Cook and Graham (1977), Mayers and Smith (1983), Schlesinger and Doherty (1985).

pay for insurance. The analysis is usually centred on the risk aversion argument, which makes possible the existence of such a mutually advantageous insurance policy. The main idea is that, since the ordinary policyholder is risk averse, he is willing to pay more than the expected loss in order to transfer his risk to the insurer. This condition makes possible the stipulation of an insurance policy, since the insurance premium required by the insurer is usually higher than the expected loss, because of the loading. An implicit assumption underlying this approach is that the policyholder deems that the insurer will be able to fulfil the obligations contracted towards him, *i.e.*, he believes that he is actually exchange an uncertain payoff (the loss he will eventually suffer) for a certain one (the insurance premium) and he is not facing any risk once he has taken out and paid for the insurance contract. However, as pointed out by Kahane *et al.*,

> *insurers are subject to default risk, and this risk (or its perception) may affect the policyholders' response to a particular premium (or loading factor) policy which is proposed by the insurer*[54].

The aim of this Chapter is to reformulate the traditional theory of the demand and supply for insurance by also taking into account the possibility of the insurer's default, *i.e.*, by dealing with insurance as an imperfect[55] tool of risk transfer.

In order to reach this purpose the rest of this Chapter is subdivided into two main sections. Section III.2 reformulates the insurance pricing theory showing that the policyholder's risk aversion explains not only the desirability of an insurance policy, even if it costs more than the expected value of losses, but also the preference for insurance policies issued by insurers who exhibit a lower risk of insolvency. The direct consequence of this approach is the acknowledgement that not only the price but also the quality of the policy (*i.e.*, the financial quality of its issuer) is relevant for the policyholder, who, therefore, exhibits a trade-off between price and quality: as the quality of the insurance policy decreases he will be willing to buy it only at a lower premium. This way of representing insurance

[54] See Kahane, Tapiero, Laurent (1989), p. 233.

demand implicitly assumes the existence of a product market discipline mechanism as defined in the previous chapters and is also consistent with the empirical evidence discussed in Chapter II.

Section III.3 aims at establishing a similar approach for the insurer's pricing approach, trying to find out a steady relationship between the optimum premium charged by the insurer and its financial quality.

III.2. The policyholder's behaviour: a formal representation

III.2.1. The standard approach

Under the simplest form, conventional expected utility theory assumes that an individual's utility, u, is a function of disposable wealth, w. Of course this assumption can be applied to the case of a policyholder. So, let's assume that the policyholder faces a possible loss due to random events that may damage his property. The distribution of the random loss L is assumed to be known (E[L] and σ are, respectively, its expected value and standard deviation). The policyholder could purchase full insurance coverage paying the premium P to the insurer, who will assume the random financial loss.

The policyholder must decide whether assuming the risk himself or transferring it to the insurer in exchange for the insurance premium. If the current wealth is w_0, the two solutions exhibit these utility values:

$$E[u(w_0 - L)] = \text{expected utility of not buying insurance}$$

$$u(w_0 - P) = \text{utility of buying insurance.}$$

[55] The "imperfect" nature of insurance arises since – *de facto* – the policyholders do not completely eliminate the risk they face, but rather they transform it into a different one, the credit

The policyholder's decision depends on the extent of P and on his preferences. Assuming that the premium P is actuarially fair, *i.e.*, P = E[L], and the policyholder exhibits a von Neumann-Morgenstern utility function $u(w)$[56], with $\frac{\partial u}{\partial w} > 0$ and $\frac{\partial^2 u}{\partial w^2} < 0$, the Jensen's inequality can be applied[57]. It follows that

$$E[u(w_0 - L)] \leq u(w_0 - E[L]). \qquad (III.1.1)$$

(III.1.1) means that if the policyholder can choose between a random payoff (whose randomness stems from the risk L) and a fixed amount equal to its expectation, he will prefer the latter, *i.e.*, the policyholders would increase his utility if he could transfer his risk to the insurer in exchange for the actuarially fair insurance premium[58]. (III.1.1) brings also to the definition of the insurance premium, say PP', which makes the policyholder indifferent between assuming the risk himself or transferring it to the insurer. PP' is the premium which satisfies the following condition:

$$E[u(w_0 - L)] = u(w_0 - PP'). \qquad (III.1.2)$$

Combining (I.2.1) and (I.2.2) we obtain:

$$u(w_0 - PP') \leq u(w_0 - E[L]). \qquad (III.1.3)$$

---

risk.

[56] See von Neumann, Morgenstern (1953). For a systematic survey on the von Neumann-Morgenstern expected utility representation, see Machina (1987).

[57] The Jensen's inequality states that for a random variable L and a twice differentiable function $u(\cdot)$, with $u(\cdot)'>0$ and $u(\cdot)''<0$,

$$E[u(L)] \leq u(E[L])$$

The economic implication of this inequality is that agents who exhibit utility functions $u(\cdot)$ with $u(\cdot)'$ and $u(\cdot)''$, *i.e.*, risk-averse agent, will evaluate an uncertain payoff at a price less than its expected value.

Since $u(w)$ is an increasing function, (III.1.3) implies that $w_0 - PP' \leq w_0 - \mathrm{E}[L]$, or $PP' \geq \mathrm{E}[L]$, or that a risk-averse policyholder is willing to pay an amount greater than the expected loss to insure a risk.

PP' is the maximal premium the policyholder is willing to pay to acquire a full insurance coverage for his risk. PP' defines the policyholder's *certainty equivalent* for that risk. The difference between the certainty equivalent and the expected value of the uncertain payoff is a measure of its *risk premium*[59]. Denoting the risk premium as RP, the following relations hold:

$$RP = PP' - \mathrm{E}[L]$$

and

$$PP' = \mathrm{E}[L] + RP \qquad \text{(III.2.1)}$$

where PP' = policyholder's certainty equivalent (*i.e.*, the largest insurance premium the policyholder would be willing to pay), $\mathrm{E}[L]$ = expected losses (as valued by the policyholder), and RP = risk premium required by the policyholder.

Having defined the notion of risk premium, (III.1.2) can be extended as follows

$$\mathrm{E}[u(w_0 - L)] = u(w_0 - PP') = u(w_0 - \mathrm{E}[L] - RP). \qquad \text{(III.1.4)}$$

These concepts are illustrated in Figure III.1.

The argument above (summarised by (III.2.1)) shows that the maximum insurance premium that an individual is willing to pay depends on two variables: (1) the losses that the individual expects to incur due to the risk he is going to transfer to the insurer and (2) the risk premium that the individual is willing to pay because of the substitution of a uncertain loss with a certain one (the insurance premium).

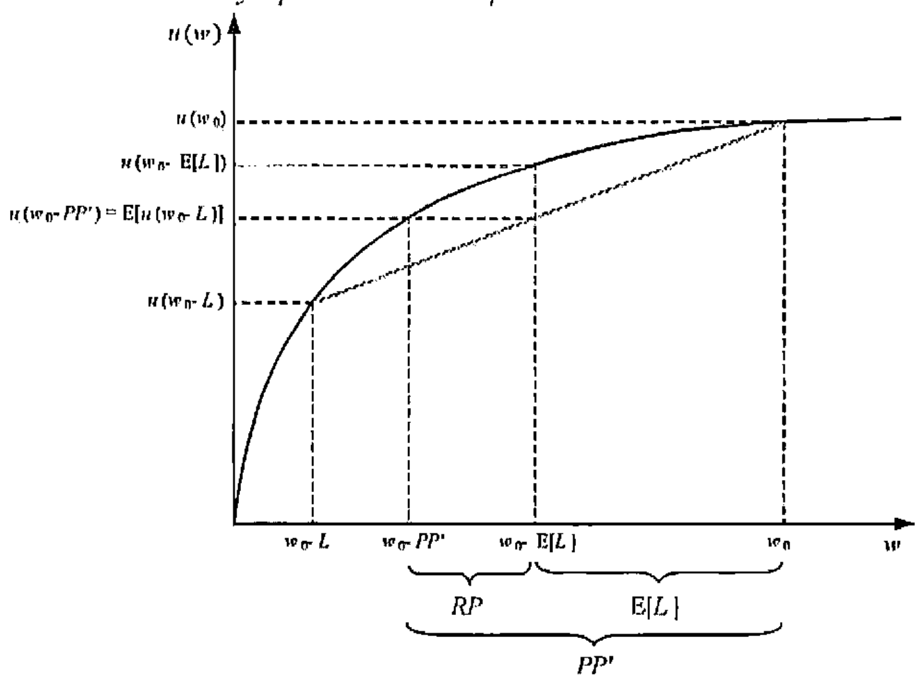
[58] The difference between the time when the premium is paid and the time when losses (and repayments from the insurer) occur is omitted. If this difference was taken into account, the expected losses should be discounted, in order to equalise sums evaluated at the same time.

[59] See Danthine, Donaldson (2002), pp. 47-50.

Attention can be drawn on the fact that *PP'* depends on both the objective characteristics of the risk ceded (as valued by the policyholder), that affect E[*L*] and (through) *RP* and the subjective policyholder's preferences (*i.e.*, his utility function), that determine (according to his risk-aversion) *RP*., *i.e.*, provided that the objective characteristics of the risk are equal, the higher the policyholder's risk-aversion, the higher the value of *RP*[60].

Figure III.1

Certainty equivalent and risk premium: an illustration



The relationship between *RP* and the individual's risk aversion can be easily shown when the risk faced by the individual-investor is a speculative one[61]. In the

---

[60] Therefore, for a risk-neutral policyholder, *RP* would be equal to zero and the *PP'* = E[*L*] would hold.

[61] See, for instance Elton, Gruber (1995), pp. 226-228, Danthine Donaldson (2002), pp. 48-49, Beltratti (2000), p. 23, where the relationship is shown expanding the expression of the individual's utility function in the point *w*+*Z* in a Taylor series around *w*, where *w* is the individual's wealth and *z* is a fair gamble with mean 0 and variance σ_z^2. The following result is derived:

$$RP \cong \frac{1}{2}\sigma_Z^2\left(-\frac{u''(w)}{u'(w)}\right) = \frac{1}{2}\sigma_Z^2 ARA(w),$$

case of a pure risk, stemming from an uncertain loss L, the assumption of an exponential utility function and a normal distribution for L[62] allows to derive a similar relationship.

The exponential utility function, $u(w) = -e^{-\alpha w}$, with $\alpha > 0$, has the following relevant features. Since $u'(w) = \alpha e^{-\alpha w} > 0$ and $u''(w) = -\alpha^2 e^{-\alpha w} < 0$, it represents the utility function for a "regular" risk averse individual, with constant absolute risk aversion, $ARA(w) = \alpha > 0$. Moreover, certainty equivalents do not depend on the wealth of the individual[63]. This statement can be verified by substituting the exponential utility function into (III.1.2), that is:

$$-e^{-\alpha(w-PP')} = \mathrm{E}[-e^{-\alpha(w-L)}],$$

which gives:

$$e^{\alpha PP'} = \mathrm{M}_L(\alpha),$$

and so:

$$PP' = \frac{\ln \mathrm{M}_L(\alpha)}{\alpha},$$

where $\mathrm{M}_L(\alpha)$ is the moment generating function of L[64].

If $L \sim \mathrm{N}(\mu, \sigma_L^2)$, the following conditions hold:

$$\mathrm{E}[L] = \mu,$$

where $ARA(w)$ is the absolute risk aversion measure, as proposed by Pratt (1964) and Arrow (1965).

[62] The assumption of a normal distribution for L is obviously a strong one, but it helps to make the derivation of the final result quite easy.

[63] Other features of the exponential utility function can be found in Gerber, Pafumi (1998), pp. 75-79.

[64] This result descends easily since for an exponential utility function the following equality $\mathrm{E}[-e^{-\alpha L}] = -\mathrm{E}[-e^{-\alpha L}] = -\mathrm{M}_L(-\alpha)$ holds.

$$M_L(\alpha) = e^{\mu\alpha + \sigma^2 \frac{\alpha^2}{2}}$$

Therefore the following expressions for PP' and RP can be obtained:

$$PP' = \frac{\ln M_L(\alpha)}{\alpha} = \frac{\ln\left(e^{\mu\alpha + \sigma^2 \frac{\alpha^2}{2}}\right)}{\alpha} = \mu + \sigma^2 \frac{\alpha}{2}$$

$$RP = PP' - E[L] = \mu + \sigma^2 \frac{\alpha}{2} - \mu = \sigma^2 \frac{\alpha}{2} = \frac{1}{2}\sigma^2 ARA(w).$$

The expression for RP shows both the dependence of RP on σ and the relationship between the risk premium and the policyholder's risk aversion. For a given expected value E[L], the higher the policyholder's risk aversion, the higher the maximum amount the policyholder would be willing to pay to avoid the risk, PP'.

III.2.2. The credit risk based approach: an introduction

(III.2.1) neglects a relevant element: the credit risk incurred by the policyholder. In point of fact, the insurance policy does not transform an uncertain loss in a certain one, but it transforms an uncertain loss in an uncertain loss of different kind. The uncertainty does not depend (solely) on the specific risk underlying the insurance policy, but also on the default risk of the insurer[65]. As pointed out by Sommer,

> *the value of an insurance policy to the policyholders is clearly dependent upon weather the insurer is solvent when the policyholders make a claim. Thus, it seems*

[65] It could be said that, in some circumstances, the default risk of the insurer is closely linked up to the specific risk underlying the insurance policy. That could especially be the case of reinsurance.

desirable that a valuation model for insurance incorporate firm default risk into the pricing of insurance[66].

Moreover, according to Phillips, Cummins and Allen,

in a competitive market with complete information, policyholders will take this limited liability position into account in deciding how much they are willing to pay for the insurance contract[67].

Following the terminology proposed by Johnson and Stulz (1987), insurance contracts can be considered *vulnerable contracts*, since their performance depends on the ability of the insurer to make the promised payment.

Hence it not really correct to compare *PP'* with the insurance premium charged by the insurer, since they are not totally homogeneous. The standard approach implicitly assumes that claims will always be paid, *i.e.*, that the probability of insurer's default is zero, while in real-world insurance markets, insurers can become insolvent, resulting in non-payment o partial payments of claims[68]. In other words, by paying the insurance premium, the policyholder would not eliminate any risk, but he would expose himself to the risk non-payment or partial payments of claims in the events of a loss on the insured risk and a total or partial default of the insurer. The insurance premium should be compared with another kind of certain equivalent which accounts also for the chance of the insurer's default.

In order to obtain this new certainty equivalent, one can assume that *PP'* is held as an uncertain amount and thus it is converted into a "true" certainty equivalent first by assessing its expected and, since policyholders are probably averse also towards the default risk of the insurer, and then by deducting a premium that accounts for this kind of risk.

In order to comprise this remark, (III.2.1) should become:

---

[66] See Sommer (1996), p. 502.

[67] See Phillips, Cummins, Allen (1998), p. 604.

[68] Cf., for instance, Cummins, Doherty, Lo (2002).

$$PP = \mathrm{E}[L] + RP - (\mathrm{E}[UL] + DRP) \qquad \text{(III.3.1)}$$

where:

- *PP* is the new[69] policyholder's certainty equivalent;
- $\mathrm{E}[UL]$ = expected uncovered losses (as valued by the policyholder)[70];
- *DRP* = risk premium required by the policyholder as regards to the risk of default of the insurer.

DPR accounts for the fact that a risk averse individual will not undertake a risk (in this case: the default risk of the insurer) unless he gets a premium higher than the expected loss. It is quite reasonable that *DPR* is linked to *PR*. Particularly, it can be thought that a policyholder who exhibits a high *PR* will also require a high *DPR*. Anyway, the actual nature of this link lies outside the aim of this analysis, therefore only a couple of remarks will be made to account for the relationship between *PR* and *DPR*.

Now, *PP* depends not only on the objective characteristics of the risk ceded (as valued by the policyholder) and on subjective policyholder's preferences, but also on the characteristics of the insurer. In fact, as the insurer's credit risk (as perceived by the policyholder) increases, the value of ($\mathrm{E}[UL] + DPR$) increases and therefore the value of *PP* decreases. This approach to the evaluation of the policyholder's certainty equivalent can be defined as a *credit risk based approach*. The effects of default risk on the demand for insurance have been incorporated into the insurance-pricing decision in many works, such as those of Doherty and Garven (1986), Johnson and Stulz (1987), Doherty (1989), Butsic (1994), Phillips, Cummins, Allen (1998), Cummins, Sommer (1996), Sommer (1996), Cummins, Danzon (1997), Cummins (1992), Babbel (1998), Downs, Sommer (1999), Cummins (2000), Girard (2002), Cummins, Nini (2002), Babbel, Gold, Merrill (2002), Cummins, Phillips (2000) and have been directly considered in the works of Tapiero *et al.* (1986), Doherty and Schlesinger (1990), and Schlesinger (2000).

[69] The word "new" refers that *PP* is derived differently from the way *PP'* has been defined and derived.

III.2.3. A comparison between the two approaches

In a world as described by (III.2.1), there will be only one certainty equivalent (PP') for each policyholder, while in a world as described by (III.3.1), the number of certainty equivalents (PP) for each policyholder will be equal to the number of the insurers that offer an insurance policy for the policyholder's specific risk.

In point of fact, using the subscript p to indicate the policyholder and the subscript i to indicate the insurer (with $p = 1, 2, \ldots, P$, and $i = 1, 2, \ldots, I$), (III.2.1) and (III.3.1) can be rewritten as follows:

$$PP'_p = E_p[L_p] + RP_p \qquad \text{(III.2.2)}$$

$$PP_{p,i} = E_p[L_p] + RP_p - (E_p[UL_i] + DRP_{p,i}) \qquad \text{(III.3.2)}$$

where the subscript p in correspondence of the $E[\cdot]$ operator means that those expectations are subjectively formulated by the p-th policyholder.

The main implication of the way to express the policyholder's certainty equivalent represented by (III.3.2) lies in the policyholder's decisional behaviour.

According to the standard approach, see (III.2.2), the policyholder's decision about the insurance purchasing[71] would rely on a simple and restrained information set. The policyholder's decisional behaviour can be summarised as follows:

> The p-th policyholder buys the policy issued by the i^*-th insurer
> if:
>
> $$IP_{p,i^*} = \min\,\{IP_{p,j}\} \quad \text{with } j \in \mathbf{J} \subset \mathbf{I}$$

---

[70] $E[UL]$ accounts for the probability of both a loss on the underlying risk and a total or partial default of the insurer.

[71] The policyholder's decision about the insurance purchasing consists of two (not necessarily separate) decisions: whether buying an insurance policy or not and, if yes, which is the most convenient policy.

where $IP_{p,j}$ is the insurance premium fixed by the j-th insurer to cover the p-th policyholder's risk, **I** is the set of all the insurers that offer a policy to cover the p-th policyholder's risk and **J** is the subset of **I** such that $IP_{p,j} \leq PP'_p$.

In this case the policyholder's decisional problem would be solved on the basis of the following information:

- the I insurance premiums ($IP_{p,i}$) set by the I insurers;
- the expected losses ($E_p[L_p]$);
- the risk premium (RP_p).

On the contrary, according to the credit risk based approach, see (III.3.2), the policyholder's decision about the insurance purchasing should rely on the following wider and more complex information set:

- the I insurance premiums ($IP_{p,i}$) set by the I insurers;
- the I expected uncovered losses ($E_p[UL_i]$);
- the expected losses ($E_p[L_p]$);
- the risk premium (RP_p);
- the I default risk premiums ($DRP_{p,i}$).

Moreover, the way this information is used is much more complex and implies a deeper specification of the policyholder's preferences, as it will be explained below.

(III.2.2) and (III.3.2) differ (a part from the mathematics) because a decisional behaviour based on (III.3.2) implies the evaluation of the insurer's creditworthiness, whereas a decisional behaviour based on (III.2.2) implies only the evaluation of the specific risk underlying the policy and the individual risk aversion.

The policyholder's certainty equivalent PP' changes according to the individual's characteristics (regarding both the risk he is transferring and his preferences). The policyholder's certainty equivalent PP, instead, may change not only according to the individual's characteristics, but also according to the insurers' characteristics[72].

Given (III.2.2), (III.3.2) could be expressed as follows:

$$PP_{p,i} = PP'_p - (E_p[UL_i] + DRP_{p,i}) \tag{III.3.3}$$

and, since $(E_p[UL_i] + DRP_{p,i}) \geq 0$, it will be $PP_{p,i} \leq PP'_p$.

The comparison between (III.2.2) and (III.3.2) requires further considerations. One of them concerns the meaning of $E_p[UL_i]$. $E_p[UL_i]$ is the policyholder's evaluation regarding the i-th insurer's default[73]. It therefore depends on the i-th insurer's reliability, which depends, in turn, on the i-th insurer's riskiness. This amount corresponds to Butsic's definition of expected policyholder deficit:

> *a reasonable measure of insolvency risk is the expected value of the difference between the amount the insurer is obligated to pay to the claimant and the actual amount paid by the insurer. We will call this difference the* policyholder deficit[74].

In addition, Butsic shows that this value is equivalent to a put option held by the insurer's owners. In effect, because liabilities may exceed the insurer's assets, its policyholders have given the insurer's owners the *option* to abandon full payment of claims.

In the option view of the firm, the value of the policyholders' claim on the firm is equal to the present value of losses minus the value of the *insolvency put option*[75]. The insolvency put option is the expected loss to policyholders due to the possibility that the firm will default.

It is clear that the value of this put option attributable to a single insurance policy depends on the features of the policy itself; in other words, different insurance

---

[72] The higher the insurer's risk of default, the lower the policyholder's insurance certainty equivalent PP'.

[73] More exactly the estimated losses depend on the probability of default of the firm, the loss rate given default and the exposure at default.

[74] See Butsic (1994), p. 660.

policies will exhibit different values of their insolvency put option, depending on both (1) the marginal contribution of the ceded risk to the insurer's overall risk, and (2) the assumption about the priority in bankruptcy of the various policies.

As far as the first aspect is concerned, it is clear that the ceded risk can affect the term σ of the insolvency put. The greater the relative weight of the ceded risk, the greater the impact, through σ, on the insolvency put.

As far as the second aspect is concerned, following Phillips, Cummins and Allen, it can be assumed that policyholders divide the assets of an insolvent insurer according to an *equal priority rule* which divides the assets of the firm among the policyholders according to the value of the liability claims (L_p) they hold against the firm[76]. Therefore, the p-th policyholder will receive a proportion $w_{L_p} = \frac{L_p}{\sum_{p=1}^{P} L_p}$ of the total assets of the firm in the event of default. Using the equal priority rule, the value of the policyholder's insolvency put option is

$$PPut_{p,i} = w_{L_p} PPut_i \qquad \text{(III.4)}$$

where:

- $PPut_{p,i}$ is the i-th insurer's insolvency put option attributable to the p-th policyholder;
- $w_{L_p} = \frac{L_p}{\sum_{p=1}^{P} L_p}$;
- $PPut_i$ is the i-th insurer's overall insolvency put option, defined as the value $Put(A, L, r, \tau, \sigma)$ of (A1) in Appendix[77].

Having defined $E_p[UL_i]$ as $PPut_{p,i}$, (III.3.3) becomes:

---

[75] Cf. Appendix for further discussion.

[76] Phillips, Cummins, Allen (1998), p. 605, make this assumption in order to allocate the cost of the insolvency put option to the different lines of business.

$$PP_{p,i} = PP'_p - (PPut_{p,i} + DRP_{p,i}). \qquad \text{(III.3.4)}$$

III.2.4. The credit risk based approach: some specifications

According to (III.3.4), the policyholder is willing to pay more to transfer a certain risk to a less risky insurer. (III.3.4) expresses, to a certain extent, the policyholder's preferences, since it depends on both PP'_p and $DRP_{p,i}$ which depend in turn, as pointed out above, on the policyholder's preferences.

Moreover, (III.3.4) defines a sort of boundary for the insurance policies that the policyholder is willing to buy. If the *i*-th insurer offered a policy with a premium $IP_{p,i}$ such that $IP_{p,i} \leq PP_{p,i}$, the policyholder would increase his utility by subscribing it[78]. Therefore, (III.3.4) defines the set of all the policies (represented as a pair $PPut_{p,i}$, $PP_{p,i}$) that the policyholder would be willing to buy. Let's name it the "*acceptable set*" for the *p*-th policyholder.

It is clear that policies with $PP_{p,i} \leq 0$ or with $PPut_{p,i} < 0$ are meaningless. Hence the acceptable set is expressed by (III.3.4), under the conditions $PP_{p,i} > 0$ and $PPut_{p,i} \geq 0$.

The acceptable set can be represented graphically in a diagram $PPut_{p,i}$, $PP_{p,i}$. To this end some specifications regarding $DRP_{p,i}$ are needed.

Example A

One of the possible ways to specify $DRP_{p,i}$ is to assume it as a multiple of $PPut_{p,i}$, *i.e.*, $DRP_{p,i} = \beta_p \, PPut_{p,i}$, where β_p is a positive constant. In this case (III.3.4) becomes (see Figure III.2):

---

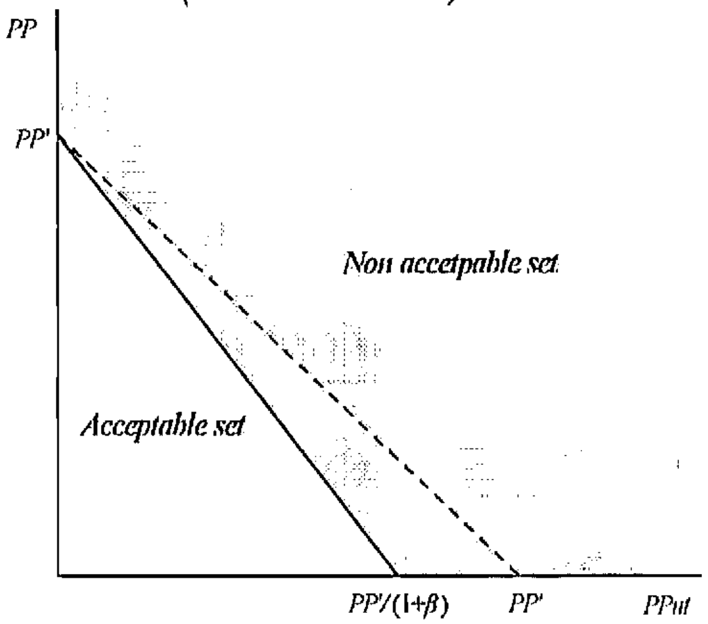
[77] Obviously, $PPut_i$ is the value of the insolvency option after the *p*-th policyholder's risk has been transferred to the *i*-th insurer.

[78] A clarification regarding the notion is needed: *PP* and *PPut* are, respectively, the insurance premium and the insolvency option premium as evaluated by the policyholder, while *IP* and *IPut* are the corresponding values as fixed, or evaluated, by the insurer. The different notation allows to take into account any difference in the way these values are estimated by the policyholder and the insurer.

$$PP_{p,i} = PP'_p - (1+\beta_p)\, PPut_{p,i} \qquad \text{(III.3.4A)}$$

Figure III.2

The acceptable set of insurance policies for a policyholder averse towards the insurer's default risk (constant risk aversion)



The meaning of this specification is clear: the higher the insolvency option, the higher the premium the policyholder will ask to undertake that risk. In this case β_p can be assumed as a parameter which summarises the policyholder aversion towards the insurer's default risk. A risk neutral policyholder (more precisely: a policyholder neutral towards the insurer's default risk) will exhibit a risk parameter $\beta_p = 0$ (see Figure III.3).

Since (III.3.4A) depends on the policyholders' risk aversion, two policyholders (denoted as policyholder *a* and policyholder *b* in the Figure III.4) exposed to the same risk (with the same values $E_p[L]$ and σ) will exhibit a different boundary (III.3.4A) if they have different degrees of risk aversion. Particularly, if $DRP_{p,i}$ and RP_p are assumed to be positively correlated, the higher the risk aversion (as policyholder *a*), the higher the value of the intercept PP'_p and the higher the slope of (III.3.4A). On the contrary, the lower the risk aversion (as policyholder *b*), the lower the value of the intercept PP'_p and the lower the slope of (III.3.4A).

Figure III.3

The acceptable set of insurance policies for a policyholder neutral towards the insurer's default risk

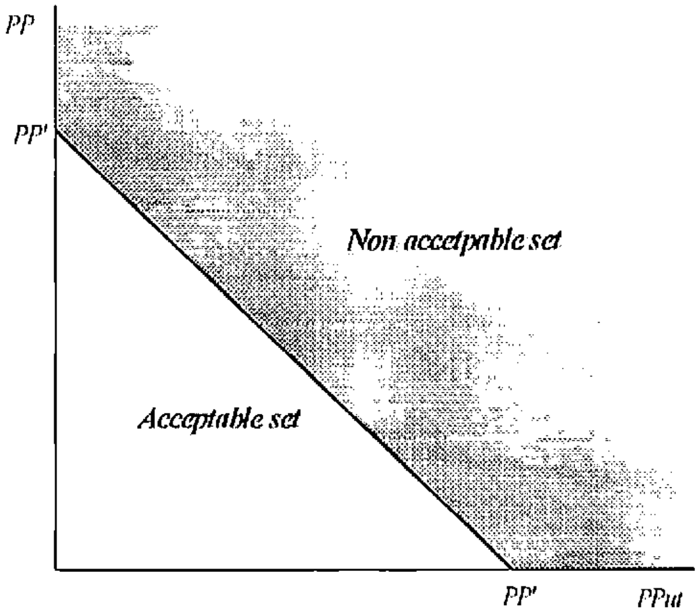
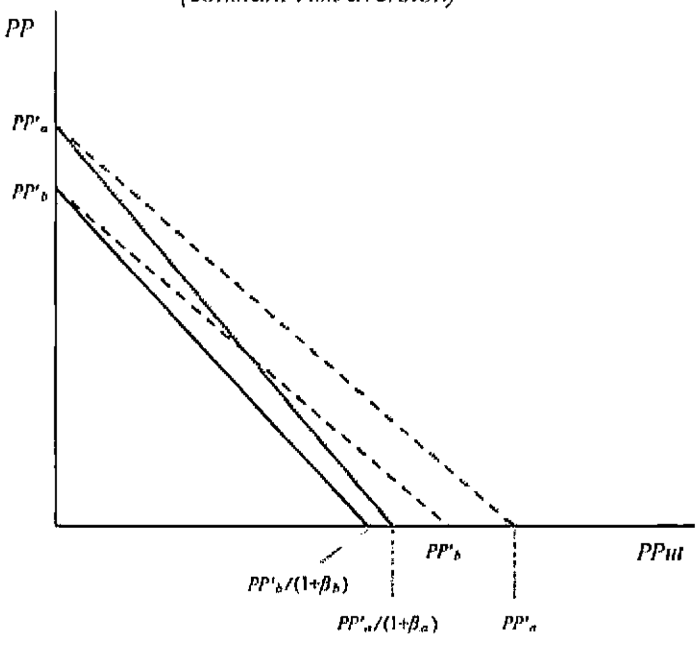


Figure III.4

The acceptable set of insurance policies for two policyholders which exhibit different aversion towards the insurer's default risk (constant risk aversion)



If the correlation between $DRP_{p,t}$ and RP_p holds, it is clear that the policyholder's risk aversion affects the set of the admissible policies in two different ways: a larger risk aversion tends to enlarge the set of the policies issued by low risk insurers and restrict the set of the policies issued by high risk insurers.

Resuming the previous example of the policyholder with an exponential utility function and facing a normally distributed uncertain loss, L, the default risk parameter β_p can be assumed to be equal to a multiple t_p (t_p>0) of the absolute risk aversion α_p, that is $\beta_p = t_p \cdot \alpha_p$. In this case (III.3.4A) becomes

$$PP_{p,t} = \mu + \sigma^2 \frac{\alpha_p}{2} - (1 + t_p \alpha_p) PPut_{p,t}. \qquad \text{(III.3.4A.1)}$$

According to this view, the policyholder's risk aversion is a complex issue, since it affects two different aspects: the aversion towards the uncertainty stemming from the specific risk underlying the insurance policy and the aversion towards uncertainty stemming from the probability of the insurer's default. Obviously, each individual will have his own way to combine these two aspects. This way reflects and expresses the policyholder's utility function and therefore his preferences. Indeed, this result is consistent with Schlesinger's following remark:

> *It also is not difficult to show that, in contrast to the case with no default risk, an increase in risk aversion will not necessarily lead to an increase in the level of insurance coverage. Although a more risk-averse individual would value the additional insurance coverage absent any default risk, higher risk aversion also makes the individual fear the worst-case outcome (a loss and an insolvent insurer) even more*[79].

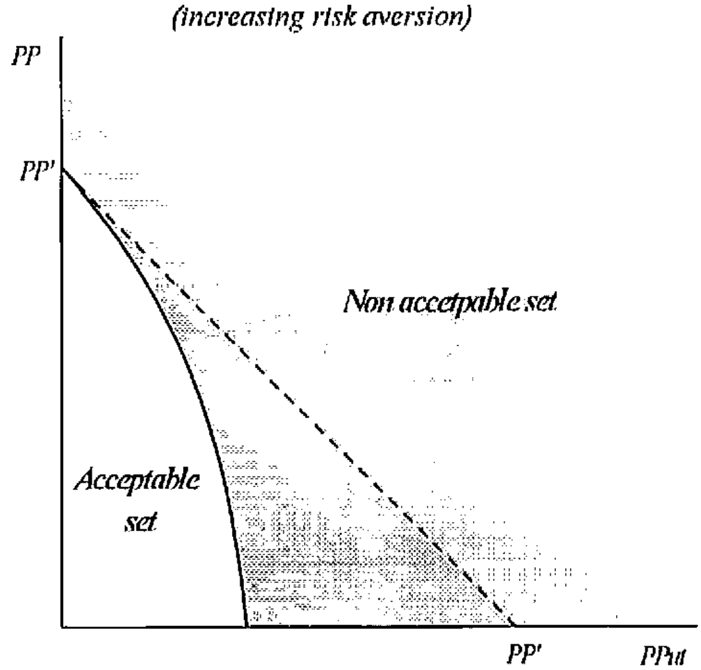
Example B

Clearly, $DRP_{p,t} = \beta_p \, PPut_{p,t}$ is just one of the possible ways to specify $DRP_{p,t}$. It assumes that the reduction of the premium asked by the policyholder following an increase in $PPut_{p,i}$ is independent of its initial value[80]. In effect, the slope of (III.3.4A), $-(1+\beta_p)$, is constant, since (III.3.4A) is the equation of a line.

Another possible way to specify $DRP_{p,i}$ could be to assume that the reduction of the premium asked by the policyholder following an increase in $PPut_{p,i}$ depends on the value of $PPut_{p,i}$ itself. It is sensible to assume that the higher the value of $PPut_{p,i}$, the higher the reduction of the premium asked by the policyholder (as depicted in Figure III.5).

Figure III.5

The acceptable set of insurance policies for a policyholder averse towards the insurer's default risk (increasing risk aversion)



One possible specification for $PP_{p,i}$ which satisfies this condition could be the following

$$PP_{p,i} = PP'_p + \gamma_p \, PPut_{p,i}^2 + \theta_p \, PPut_{p,i} \qquad \text{(III.3.4B)}$$

where $\gamma_p < 0$ and $\theta_p < -1$[81].

[79] See Schlesinger (2000), p. 145.

[80] *I.e*, the reduction is always $(1+\beta_p)$ times the increase in $PPut_{p,i}$.

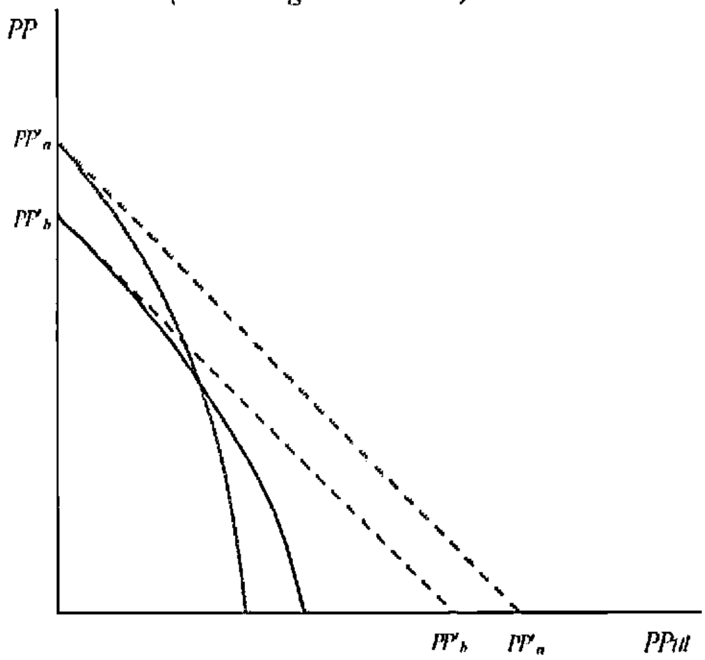
[81] The restrictions over γ_p and θ_p cause (III.3.4B) to represent the equation of a parabola passing through the point (0, PP'_p) and laying below the equation $PP_{p,i} = PP'_p - PPut_{p,i}$. The latter

γ_p and θ_p are negatively linked to PP'_p, that is if PP'_p increases (decreases) γ_p and θ_p decrease (increase)[82]. If the policyholder has an exponential utility function, one can assume that his absolute risk aversion parameter, α_p, is negatively related to/with γ_p and/or θ_p, *i.e.*, $DRP_{p,i}$ is positively linked to a RP_p.

Again, since (III.3.4B) depends on the policyholders' risk aversion, two policyholders (denoted as policyholder *a* and policyholder *b* in the Figure III.6) exposed to the same risk (with the same values $E_p[L]$ and σ) will exhibit a different boundary (III.3.4B) if they have different degrees of risk aversion. Particularly, if $DRP_{p,i}$ and RP_p are assumed to be positively correlated, the higher the risk aversion (as the policyholder *a*), the higher the value of the intercept PP_p' and the higher the slope of (III.3.4B). On the contrary, the lower the risk aversion (as the policyholder *b*), the lower the value of the intercept PP_p' and the lower the slope of (III.3.4B).

Figure III.6

The acceptable set of insurance policies for two policyholders which exhibit different aversions towards the insurer's default risk (increasing risk aversion)



restriction is needed to obtain an acceptable set consistent with the hypothesis of the policyholders' aversion towards the insurer's default risk.

82 The way γ_p and θ_p react to PP'_p can be very different according to the individual's characteristics. One can assume that variations of PP'_p bring about changes only in γ_p, or only in θ_p, or both.

It can be noted that if policyholders had the same risk, *i.e.*, if all policyholders exhibit the same value of $E_p[L]$ and σ, the previous analysis could be carried out by measuring on the vertical axis just the risk premium RP_p instead of $PP_{p,i}$. In fact the values of $PP_{p,i}$ pointed out in the previous analysis do not reflect only the differences in risk aversion, but also any difference in the expected value of the loss.

The definition (III.3.4) deserves a further explanation. (III.3.4) is the set of the points (*i.e.*, the set of the pairs $PPut_{p,i}$, $PP_{p,i}$) where the p-th policyholder is indifferent between incurring a risk with expected loss $E_p[L]$ and standard deviation σ or buying an insurance policy from the i-th insurer (which exhibits a certain default risk), paying a premium $PP_{p,i}$. It means that each of the points of (III.3.4) exhibits the same utility, hence (III.3.4) is an indifference curve.

On the contrary, any point below (III.3.4) corresponds to a policy which would increase the policyholder's utility. Let's consider, for instance, the point with coordinates (0, $PP'_p - \varepsilon$), where ε is a small positive amount and let's compare it with any point on (III.3.4). If the policyholder could buy such a policy, he would be better off. In fact he prefers it to the point with coordinates (0, PP'_p) which belongs to (III.3.4), since he will be exposed to the same default risk (they have the same value of $PPut$), but he will pay less. Moreover, according to his preferences, the policyholder would be indifferent between this policy and some others which imply certain degrees of default risk, but are adequately cheaper. If the kind of relationship which makes the policyholder indifferent between a policy with a zero default risk and risk ones was always the same, the indifference curves would exhibit the same shape (as depicted in Figure III.7). Finally, the lower the indifference curve the higher the corresponding policyholder's utility, with the "bliss point" laying where both $PPut_{p,i}$ and $PP_{p,i}$ are equal to zero (such a policy would be sold free and its issuer would not exhibit any default risk).

Having defined the policyholder's indifference curves, the policyholder's decisional problem can be summarised as follows.

The p-th policyholder buys the policy issued by the i^*-th insurer if:

$$PP_{p,i^*} - IP_{p,i^*} = \max \{PP_{p,j} - IP_{p,j}\} \text{ with } j \in \mathbf{J} \subset \mathbf{I}$$

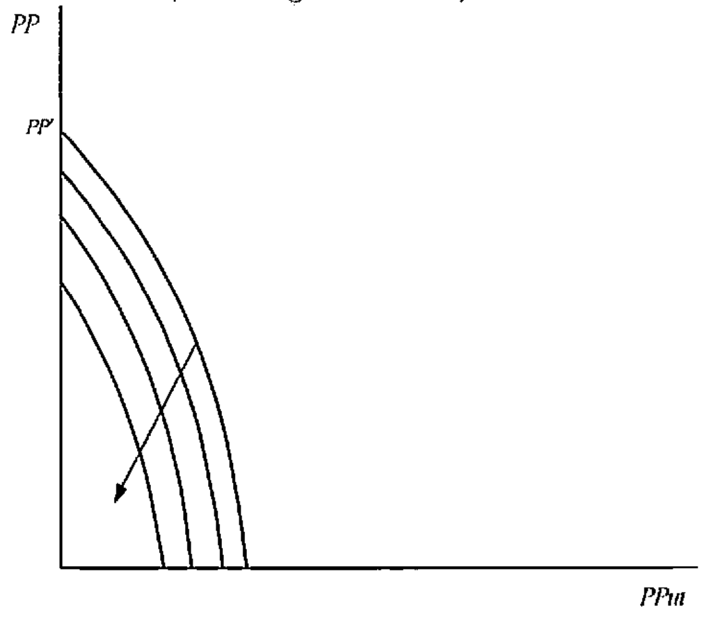
where $IP_{p,j}$ is the insurance premium fixed by the j-th insurer to cover the p-th policyholder's risk, $PP_{p,j}$ is the p-th policyholder's certainty equivalent corresponding to the j-th insurer, $\mathbf{I}$ is the set of all the insurers that offer a policy to cover the p-th policyholder's risk and $\mathbf{J}$ is the subset of $\mathbf{I}$ such that:

$$IP_{p,j} \leq PP'_p - (PPut_{p,i} + DRP_{p,i})$$

(*i.e.*, the premiums charged by the insurers belong to the admissible set).

Figure III.7

The indifference curves for a policyholder which exhibit aversion towards the insurer's default risk (increasing risk aversion)

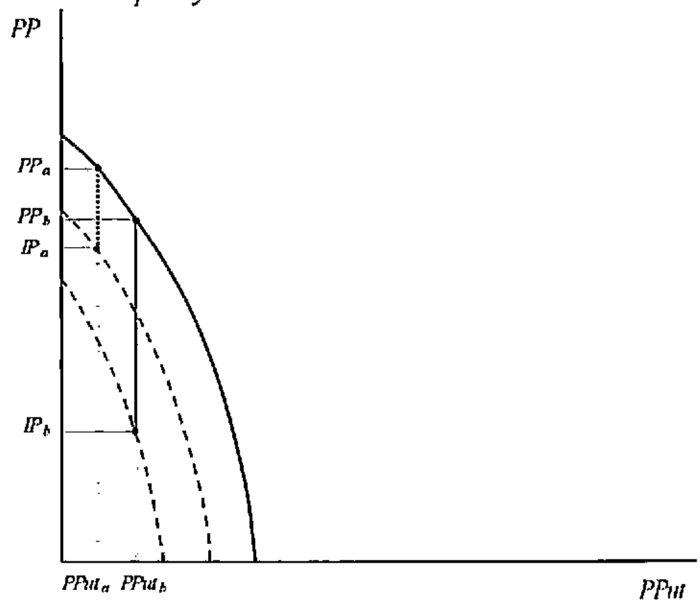


In other words, it could be thought that, given some insurance policies (belonging to the acceptable set), a policyholder whose preferences are defined according to

(III.3.4), will select the one that exhibits the greatest difference between the premium charged by the insurer ($IP_{p,i}$) and the policyholder's certainty equivalent ($PP_{p,i}$) corresponding to that policy's insolvency option, $PPut_{p,i}$. In a sense, the policyholder's certainty equivalent corresponding to each policy's insolvency option, $PPut_{p,i}$, represents a more complete and correct[83] definition of certainty equivalent, hence a comparison between the differences $PP_{p,j} - IP_{p,j}$ is consistent, since it is nothing else than a comparison between certain values or certainty equivalents. Figure III.8 shows this argument in the case of a selection between two insurance policies, *a* and *b*. In this case, policy *b* is preferred since it implies a *price-insolvency risk* mix better lined up with the policyholder's preferences.

Figure III.8

The policyholder's decisional behaviour



The main indications which can be drawn from the previous analysis are the following:

[83] In fact it takes into account not only the uncertainty related to the underlying risk but also the uncertainty related to the insurer's default risk.

1) a negative relationship exists between the insurers' default risk and the premium that the policyholder will be willing to pay for their insurance coverage commitments[84];
2) the higher the policyholder's risk aversion, the more emphasised is this relationship;
3) each policyholder exhibits a maximum value for the insurer' s default risk he is willing to assume by underwriting the policy he has issued;
4) the higher the policyholder's risk aversion, the lower this maximum value;
5) the higher the policyholder's risk aversion, the higher the premium he will be willing to pay for a default risk free insurance policy and the lower the premium he will be willing to pay for a risky insurance policy.

Some of these indications can be directly used to try to develop a more complete theory of insurance demand. Summing up, insurance demand depends, negatively, on two elements: the premium charged by the insurer and his default risk or "financial quality".

This definition of insurance demand is subject to the obvious condition that the policyholder is able to evaluate the insurer's default risk. On the contrary, where this condition did not hold, the demand for insurance would depend only on the premium charged by the insurer and competition between insurers would be played with regard only to the price. In a world of well-informed risk averse policyholders, a boundary for the insurer's default risk would exist in addition to a boundary for the premium. Consequently, the insurers would be aimed by the purpose of offering an insurance policy that satisfies the policyholders' preferences.

[84] This result is consistent with the observation suggested by Kahane, Tapiero, Laurent (1989), p. 234. According to them:

> *The insureds are subject to conflicting interests: while they are interested in the stability of their insurer, they are less happy about the higher loading factor and higher premiums that this stability implies.*

This result is also consistent with the empirical evidence cited in Chapter II.

III.3. The insurer's behaviour

A further stage of the analysis requires a specification of the insurer's behaviour and particularly the way he sets his insurance premiums, $IP_{p,i}$.
Obviously, the following general condition must hold:

$$IP_{p,i} \geq E_i[L_p] \qquad \text{(III.5.1)}$$

that is: the insurer can not charge a premium lower than the expected loss on the insurance policy. The difference between $IP_{p,i}$ and $E_i[L_p]$ is called *Loading*, *i.e.*:

$$IP_{p,i} = E_i[L_p] + Loading_{p,i}. \qquad \text{(III.5.2)}$$

Since $E_i[L_p]$ is unequivocally defined according to the insurer's evaluation of the policyholder's loss distribution, the most difficult part in the premium's evaluation is the definition of *Loading*'s determinants. Actually, *Loading* can be decomposed into two components: one accounts for the expenses of doing business and the other depends on the insurer's risk exposure. Naming these components, respectively, $Loading(E)_{p,i}$, and $Loading(R)_{p,i}$, (III.5.2) becomes:

$$IP_{p,i} = E_i[L_p] + Loading(E)_{p,i} + Loading(R)_{p,i}. \qquad \text{(III.5.3)}$$

In fact, the insurance premium is usually expressed as the sum of the following elements: *pure premium*, *expenses of doing business* and *risk charge*[85].
The pure premium is the average amount of loss, and hence is equal to $E_i[L_p]$.
The expenses of doing business include commissions to agents or brokers, salaries to employees, rental of building space or equipment, supplies, taxes, and so on. Typically, expenses can be measured as a percentage of the pure premium, or as a flat amount per contract. In those cases it will be, respectively, $Loading(E)_{p,i}$, $= l\cdot E_i[L_p]$, where $0 < l < 1$ and $Loading(E)_{p,i}$, $= L$, where L is a fixed amount of money. It is clear that $Loading(E)_{p,i}$ (*i.e.*, l or L) depends on the cost structure of

[85] See, for instance, Hogg, Klugman (1984), pp. 3-4.

the insurer: the higher the expenses of doing business, the higher *Loading(E)*. This expense-driven component of *Loading* is not affected by the insurer's risk exposure.

The risk charge, $Loading(R)_{p,t}$, is the payment to which the insurer is entitled for exposing its surplus, and hence its equity holders, to risk fluctuations. All things being equal, the risk charge increases when either the amount of equity to be rewarded[86] or the likelihood of the inadequacy of policyholders' funds to meet the insurer's liabilities increases. According to the previous specification, $Loading(R)_{p,t}$ depends basically on two variables, E_t (*i.e.*, A_t/L_t) and σ_t, so that (III.5.3) can be rewritten as

$$IP_{p,t} = E_t[L_p] + Loading(E)_{p,t} + Loading(R_\sigma, R_{A/L})_{p,t} \qquad \text{(III.5.4)}$$

where the argument R_σ in $Loading(R_\sigma, R_{A/L})_{p,t}$ refers to the premium loading set to cover the risk likelihood of an inadequacy of policyholders' funds and the argument $R_{A/L}$ refers to the premium loading set to reward the equity holders and generate adequate profits.

It can be assumed that $Loading(R)_{p,t} = Loading(R_\sigma, R_{A/L})_{p,t}$ depends on σ_t and A_t/L_t as follows:

$$\frac{\partial Loading(R)_{p,t}}{\partial \sigma_t} > 0$$

and

$$\frac{\partial Loading(R)_{p,t}}{\partial (A_t/L_t)} > 0 .$$

[86] Holding equity capital in an insurance company is costly due to regulatory costs, agency costs from unresolved owner-manager and owner-policyholder conflicts, the costs of adverse selection and moral hazard in insurance underwriting and claims settlements, corporate taxation and other market frictions. For further discussion, see Cummins and Nini (1994).

As the insurer's risk parameter σ_i increases, it becomes less likely that the pure premiums (increased by the earnings descending from the investment of the premiums) will be able to meet the insurer's liabilities, forcing the insurer to increase the loading, so $\frac{\partial Loading(R)_{p,i}}{\partial \sigma_i} > 0$.

On the other hand, as the ratio A_i/L_i, increases as a result of an increase in surplus E_i, costs arising from using equity capital cushions also increase, so $\frac{\partial Loading(R)_{p,i}}{\partial (A_i/L_i)} > 0$.

As stated earlier, σ_i and A_i/L_i are the two endogenous variables that affect the value of the insolvency put, $IPut_{p,i}$[87]. Since the put value varies inversely with the asset-to-liability ratio and directly with the risk parameter, the relationship between the optimum premium charged by the insurer, $IP_{p,i}$, and the put value, $IPut_{p,i}$, is not unambiguous, because it depends on the mix between σ_i and A_i/L_i underlying a specific value of $IPut_{p,i}$.

The only useful remark that can be made as regard the relationship between $IP_{p,i}$ and $IPut_{p,i}$ concerns the likely path of $IP_{p,i}$ as a result of a change in σ_i or A_i/L_i (and, consequently, in $IPut_{p,i}$).

In order to discuss in depth this relationship, it could be useful to derive an expression similar to (III.3.3). Splitting up the terms which depend on the insurer's default risk and those which do not depend on it and calling $IP'_{p,i} = E_i[L_p] + Loading(E)_{p,i}$ the term which does not depend on the risk of default of the insurer, (III.5.3) becomes:

$$IP_{p,i} = IP'_{p,i} + Loading(IPut)_{p,i} \tag{III.5.5}$$

[87] The notation $IPut_{p,i}$, instead of $PPut_{p,i}$ is used here to stress the fact that we refer to the insolvency option as evaluated by the insurer (which may differ from the analogous option if evaluated by the policyholder).

where $Loading(IPut)_{p,i}$ is placed instead of $Loading(R)_{p,i}$ since this risk charge depends on the two endogenous variables that affect the value of the insolvency put.

This latter equivalence allows to rewrite (III.5.4) as

$$IP_{p,i} = IP'_{p,i} + Loading(IPut_{\sigma}, IPut_{A/L})_{p,i}. \qquad \text{(III.5.6)}$$

However, $IP'_{p,i}$ does not define clearly $IP_{p,i}$ when $IPut_{p,i}$ is worth zero. In fact, a zero value for $IPut_{p,i}$ could be obtained with different mixes of σ_i and A_i/L_i. For instance, a zero $IPut_{p,i}$ could be achieved by increasing indefinitely the surplus cushion E_i (and the ratio A_i/L_i) or by decreasing the intrinsic riskiness σ_i. In these two cases the corresponding value of $IP_{p,i}$ could not be the same, but $IP_{p,i}$ would probably be higher if zero $IPut_{p,i}$ were achieved by increasing the surplus cushion E_i, instead of by decreasing the intrinsic risk σ_i.

According to this remark, it can be said that even if $IP'_{p,i}$ does not depend on $IPut_{p,i}$, it does not settle the intercept, which depends also on $IPut_{p,i}$. Let's call this intercept $IP''_{p,i}$. $IP''_{p,i}$ depends on $IPut_{p,i}$, since its value varies according to the mix of σ_i and A_i/L_i that makes $IPut_{p,i}$ equal to zero.

Summing up, each insurer exhibits its own $IP''_{p,i}$ which depends[88] on his specific cost structure (which determines $Loading(E)_{p,i}$) and also his specific combination of σ_i and A_i/L_i.

As well as $IP''_{p,i}$, the way $IP_{p,i}$ changes further to an increase in $IPut_{p,i}$ is not unambiguous. Again, it depends on the source of the increase in $IPut_{p,i}$. If $IPut_{p,i}$ increases because of an increase in σ_i, $IP_{p,i}$ will increase, since the $\dfrac{\partial Loading(IPut_{\sigma}, IPut_{A/L})_{p,i}}{\partial \sigma_i} > 0$ effect prevails. On the other hand, if $IPut_{p,i}$ increases because of a decrease in E_i (*i.e.*, in A_i/L_i), $IP_{p,i}$ will decrease, since the $\dfrac{\partial Loading(IPut_{\sigma}, IPut_{A/L})_{p,i}}{\partial (A_i/L_i)} > 0$ effect prevails (see Figure III.9[89]).

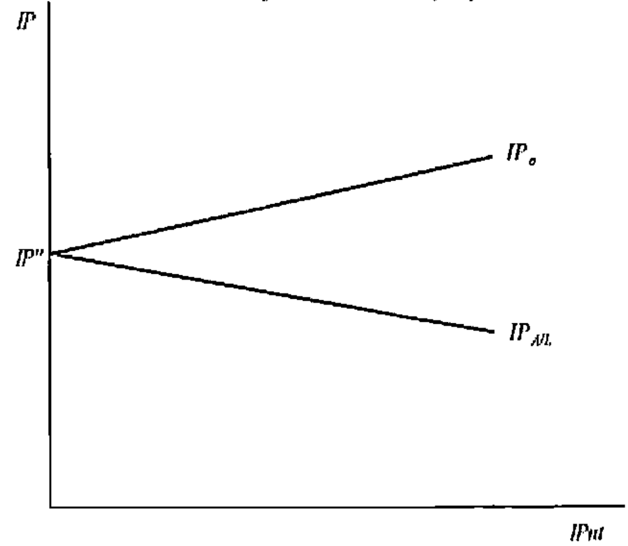
---

[88] Apart from the characteristics of the risk insured, measured by $E_i[L_p]$.

[89] The relationships between $IPut_\sigma$ or $IPut_{A/L}$ and IP depicted in Figure III.9 are linear, but they could exhibit more complicated shapes. In this analysis some aspects of those relationships are

Figure III.9

The premium charged by the insurer as a function of the determinants of the insolvency option



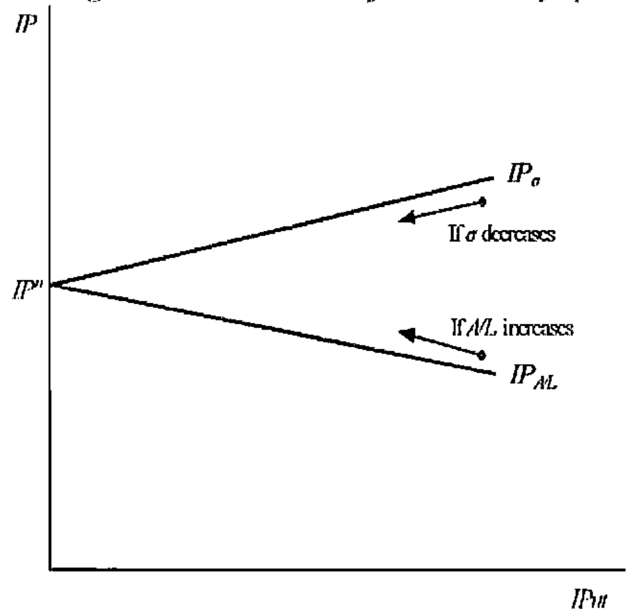
The meaning of this remark is clear: given a value of $IPut_{p,t}$, the corresponding value[90] of $IP_{p,t}$ can not be determined without an accurate specification of the mix of σ_t and A_t/L_t which underlies $IPut_{p,t}$. Anyway, given a stating value of $IPut_{p,t}$ (whatever path it comes from), the previous considerations allow to determine how the value of $IP_{p,t}$ should change as response to a change in σ_t or A_t/L_t (see Figure III.10).

The meaning of (III.5.6) deserves a further specification. One way to clarify the meaning of (III.5.6) could be the following: an insurer who fixed the premium according to (III.5.6) would be covering the expenses and remunerating adequately his production factors, since he would be getting from the policyholders a premium large enough to cover the expenses and the risk he is incurring and to remunerate the equity holders.

omitted. For instance, we do not make allowance for the fact that (the amount of equity being equal) any change of the intrinsic riskiness, σ, could actually affect the cost of equity.

Figure III.10

The path of the premium charged by the insurer as function of the change in the determinants of the insolvency option



A premium higher than the value rising out of (III.5.6) would grant a rent to the insurer and, in a sense, the resulting extra profits would increase the insurer's solvency. On the contrary, a premium lower than the value rising out of (III.5.6) would deteriorate the insurer's economic and financial equilibrium, since he would find harder to meet the claim liabilities and/or remunerate adequately the equity holders, and, as a direct consequence, the insurer's solvency would decrease. It is clear that any increase/decrease in the insurer's solvency driven by a higher/lower premium depends mainly on how generalised is the insurer's attitude to set the insurance premiums in that way[91].

Therefore (III.5.6) represents a sort of steady-state condition, that is a set of combinations of risk levels and premiums that prevents any change in the default risk of the insurer. In other words, for a given mix σ_t^* and A_t^*/L_t^* and the related values $IPut_{p,t}^*$ and $IP_{p,t}^*$ (*i.e.*, respectively, the value of an insolvency put when $\sigma_t = \sigma_t^*$ and $A_t/L_t = A_t^*/L_t^*$ and the value which satisfies (III.5.6) when $IPut_{p,t} = IPut_{p,t}^*$), (III.5.6) explains how $IP_{p,t}$ changes as a result of a change in σ_t or A_t/L_t

[90] According to (I.6.6).

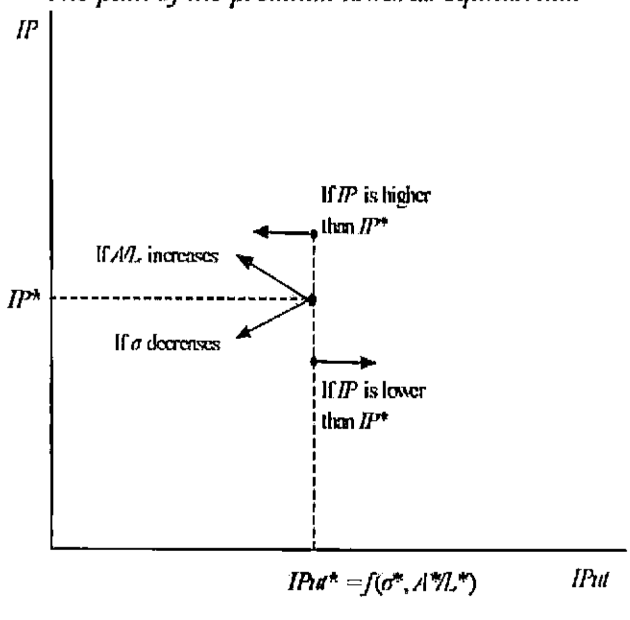
[91] *I.e.*, if the insurer usually set premiums higher than the values rising out of (I.6.6) then it riskiness could be thought to decrease (at least in the long run). A similar conclusion could not be inferred if this way of setting premiums happened just sometimes.

and how $IPut_{p,t}$ changes as a result of a disequilibrium value of $IP_{p,t}$ (see Figure III.11).

Nevertheless, in practice, there could be circumstances where the premium is far from the equilibrium premium and, indeed, disequilibria are by far more frequent. In fact, if it is possible (*i.e.*, if policyholders find it convenient), an insurer will act in order to exploit as much as possible any rent opportunity, by charging a premium higher than the value rising out of (III.5.6). Similarly, an insurer could find convenient to charge a premium lower than the value rising out of (III.5.6) if it can help to attract new policyholders.

Figure III.11

The path of the premium towards equilibrium



CHAPTER IV

Insurers' behaviour and supervisory implications in a market discipline framework

IV.1. Introduction

The analysis developed in Chapter III has a relevant implication for both the insurers' and supervisors' behaviour.

As far as insurers are concerned, the main obvious implication is that competition between insurers is played along two main dimensions: the final premium charged and their financial quality. In other words, an insurer may attempt to gain a further market demand by decreasing the premium charged and/or by improving his financial quality as perceived by policyholders. This market-driven behaviour is the ultimate effect of the product market discipline operating in the insurance sector. The actual features of such a market discipline mechanism will be described in Section IV.2. First, it will be assumed a full observability of insurers' riskiness, and then some considerations about a more realistic condition of opaqueness will be made.

Necessarily, the extent to which market discipline works affects insurance supervisory. As far as supervisory is concerned, a deeper analysis will be developed in Section IV.3.

IV.2. Some insights about the strategic interaction between policyholders and insurers

The acknowledgement that the efficiency of a contract is diminished if it is exposed to the firm's default should have remarkable impact on the firm's behaviour. In such a case, the businesses of a firm that are significantly affected by changes in customer perception of the credit standing of the firm are described as "credit-sensitive". Insurance is therefore a credit-sensitive activity.

In order to see how the acknowledgement of the credit-sensitive nature of the insurance exerts influence on the insurer's behaviour an analysis of the policyholders and insurers strategic interaction is needed.

IV.2.1. Perfect transparency

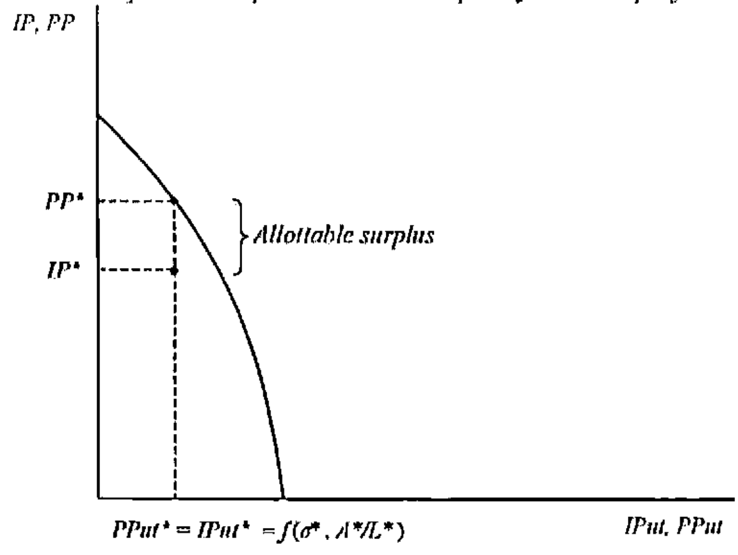
As a starting point, one can assume an interaction restricted to the p-th policyholder, whose acceptable set of policies is defined according to (III.3.4), and the i-th insurer, who offers a policy which exhibits a certain risk $IPut_{p,i}^*$ (descending from a certain mix σ_i^* and A_i^*/L_i^*) corresponding to a certain equilibrium premium $IPut_{p,i}^*$ defined according to (III.5.6). On the basis of (III.3.4), the certainty equivalent $PP_{p,i}^*$ for the p-th policyholder (corresponding to the insolvency option $IPut_{p,i}^*$) can be easily derived. Evidently, we are also assuming that $PPut_{p,i}^* = IPut_{p,i}^*$, meaning that policyholders can exactly evaluate the insolvency option. Obviously it is a brave assumption and one of the next purposes is to see the main implications which follow the relaxing of this hypothesis.

If the condition $IP_{p,i}^* < PP_{p,i}^*$ held, the policyholder would be willing to buy the policy and he will increase his utility as high as the difference $PP_{p,i}^* - IP_{p,i}^*$. In this case this difference would be a sort of surplus, entirely enjoyed by the policyholder.

Whichever premium $\overline{IP}_{p,i}$ the insurer set (provided that $IP_{p,i}^* \leq \overline{IP}_{p,i} < PP_{p,i}^*$), the policyholder would be willing to buy the policy. In this case the surplus $PP_{p,i}^* - IP_{p,i}^*$ would be allocated between the two parts. Hence the difference $PP_{p,i}^* - IP_{p,i}^*$ represent the policy's allottable surplus (as represented in Figure IV.1).

Figure IV.1

The insurer's equilibrium premium and the policyholder's preferences

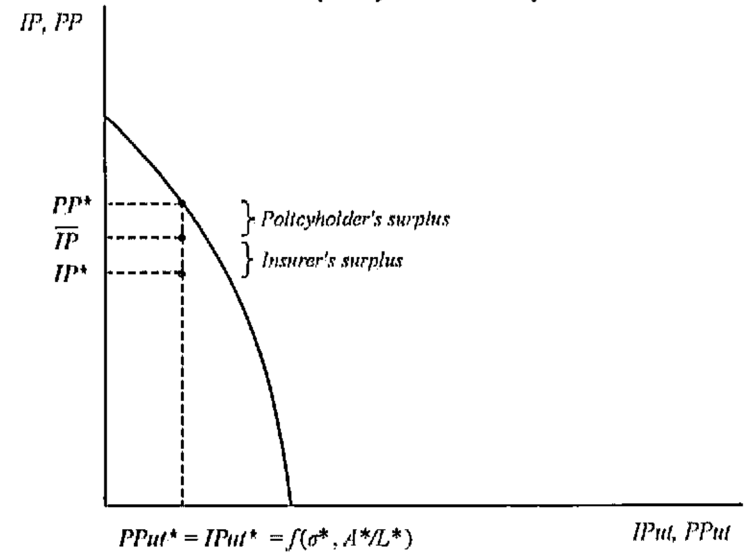


The higher $\overline{IP}_{p,i}$, the higher the portion of the surplus assigned to the insurer and the lower the portion allocated to the policyholder (see Figure IV.2).

Combining these remarks with the considerations concerning the equilibrium premium set by the insurer, another indication can be drawn: a situation where the insurer enjoys some of the allottable surplus could be beneficial to the policyholder in the following way: a premium higher than the equilibrium premium decreases the insurer's default risk, increasing, in turn, the policyholder's utility.

Figure IV.2

Insurer's and policyholder's surplus



This argument could be summarised as follows: if the policyholder is said to be concerned not only with the (absolute) insurance premium but also with the insurer's insolvency, the effects of any increase in the premium charged by the insurer are twofold: in fact a risk effect (*i.e.*, the increase in utility which follows the decrease of the insurer's default risk driven by the higher premium) joins the more common price effect (*i.e.*, the decrease in utility which follows any increase in the premium). Obviously, the relative weight of those two effects depends on either the shape of the policyholder's utility function and the decrease/increase in the insurer's default risk which stems from a premium higher/lower than the equilibrium premium. Indeed, this argument is also described by Kahane *et al.*, when they assert that

> *the default risk becomes dependent on the premiums, in addition to other factors – such as the insurer's capitalisation, the returns on investments, and the distributional properties of the loss severity and frequency. Therefore, a premium or loading factor policy must necessarily be a function of the policyholder's reactions towards the probability of insurer's default. In other words, the optimal*

premium rate depends on the default risk – which, in turn, depends on the premium rate itself[92].

If the condition $IP_{p,t}^* > PP_{p,t}^*$ held, the policy would belong to the *non acceptable set* and the policyholder would not be willing to buy it. A sort of rationing of the insurance policies would be at work, so that policies with a high premium or with a high insurer's default risk would not be subscribed. In such a case the insurer would have two alternatives: (1) fixing a premium lower than $PP_{p,t}^*$ (and therefore lower than $IP_{p,t}^*$) or (2) carrying out a conduct directed towards a decrease of the risk of default. The first alternative is clearly inconsistent (unless it s assumed that the following relationship holds: a lower premium attracts a larger number of policyholders, increases the size of the insurer's portfolio of risk, decreasing his actuarial risk and hence his default option), because it implies the insurer' s inability to remunerate adequately his production factors. The second alternative may be performed in two different ways: either by decreasing the insurer's intrinsic riskiness, σ_t, or by increasing the surplus cushion, E_t (and therefore the ratio A_t/L_t). In any case, the existence of an acceptable set depending also on the insurer's risk of default would lead insurers to adopt risk-lowering measures. A market discipline mechanism would be at work, through two possible effects: a price effect and a quantity effect. The quantity effect consists in the shifting in the insurance demand towards more safe insurers, while the price effect is the premium reduction asked by policyholders to compensate an insurer's high risk of default. The two effects are often related, since is the quantity effect which usually prompts the price effect. The market discipline mechanism could cause the insurer's exit from the market if he was not in the position to reduce the premium. As pointed out by Harrington (2004*a*), in a market discipline framework

> *deterioration in an insurer's financial condition and an attendant increase in its insolvency risk should reduce new and renewal business as policyholders gravitate to higher quality insurers.*
>
> *(...)*

[92] See Kahane, Tapiero, Laurent (1989), p. 233.

Holding prices fixed, a reduction in the volume of coverage sold will reduce revenues. Because higher insolvency risk lowers the prices that risk-sensitive policyholders are willing to pay, the insurer's average price per unit of coverage will also decline, further reducing revenues.

Moreover,

risk-sensitive demand encourages firms to commit capital ex ante to bond their promises to pay claims and thus receive higher prices. The possible loss of valuable repeat business amplifies that incentive and also discourages ex post increases in risk.

The market discipline effect would mainly lie in forcing the insurer to limit the extent of his insolvency option. Any reduction in the value of the insolvency option affects the insurer in two closely connected ways: (1) by increasing the probability of offering an insurance policy belonging to the policyholder's acceptable set; (2) by increasing the probability of exploiting any (higher) rent opportunity.

The first effect can be easily explained: because the policyholder's demand for an insurance policy is negatively related both to the price and to the insurer's financial quality (as expressed by the insolvency option implicitly ceded by the policyholders) any reduction of the insolvency option premium will increase the probability that, other things being equal, the policyholder will be willing to buy his policies.

The second effect requires a further specification: the maximum value of the insurer's rent is $PP_{p,i}^{*} - IP_{p,i}^{*}$. In any case, this difference can not exceed the value $PP_{p,i}^{*}$. Given the shape of the acceptable set (negatively related to $PPut_{p,i}$), it follows that the lower $PPut_{p,i}$, the higher the probability to exploit a (higher) rent.

The market discipline effect would also lie in forcing the insurer to choose an appropriate way to limit the extent of his insolvency option: while both a reduction of σ and an increase of E concur to decrease $IPut_{p,i}$, a reduction of σ enables also to reduce $IP_{p,i}^{*}$ increasing both the probability of offering an insurance policy belonging to the policyholder's acceptable set the probability of exploiting any (higher) rent opportunity.

From the insurer's point of view the market discipline would force the insurer to decrease his insolvency risk by decreasing his intrinsic riskiness.

The next step in the analysis of the strategic interaction between policyholders and insurers is to assume the existence of many insurers, competing to serve the policyholder. Let's assume that two insurers (say *a* and *b*) offer an insurance contract to the *p*-th policyholder. The two insurers could exhibit different equilibrium premiums, $IP_{p,a}^*$ and $IP_{p,b}^*$, since they could have, for instance, different cost structures (hence different values of $Loading(E)_{p,i}$) or different risk exposures (hence different values of $Loading(IPut)_{p,i}$). The policyholder would choose the policy issued by the i^*-th insurer which maximises his utility, *i.e.*, for which the difference between the insurance premium set by the insurer ($IP_{p,i}$) and the corresponding policyholder's certainty equivalent ($PP_{p,i}$) is as high as possible.

Let's assume that the equilibrium insurance premiums and the insolvency option premiums of the two insurers are those depicted in Figure IV.3. If the insurers set their respective equilibrium premiums, the policyholder's demand would be entirely captured by insurer *b*. In fact insurer *b* benefits from a competitive advantage over insurer *b*.

The extent of this advantage depends not only on the insurers' cost structures and risk exposures, but also on the features of the policyholders' preferences. Figure IV.4 shows how the competitive advantage changes if the policyholders' preferences change as well even though the insurers are the same. In fact, in the left-side diagram it is insurer *b* who exhibits a competitive advantage, while the insurer *a* enjoys a competitive advantage if policyholders' preferences are like those depicted in the right-side diagram.

Insurer *b* in Figure IV.3 could monetise its competitive advantage by increasing the premium by an amount ε_b, asking a premium $\overline{IP}_{p,b}$ ($= IP_{p,b}^* + \varepsilon_b$) such that the policyholder will continue to prefer it to insurer *a*. In this way insurer *b* would enjoy the rent ε_b.

Figure IV.3

Competition between insurers a *and* b

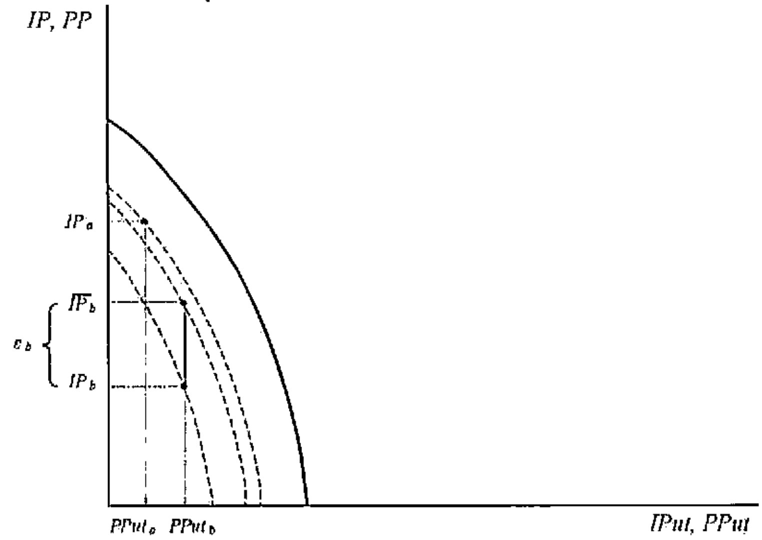
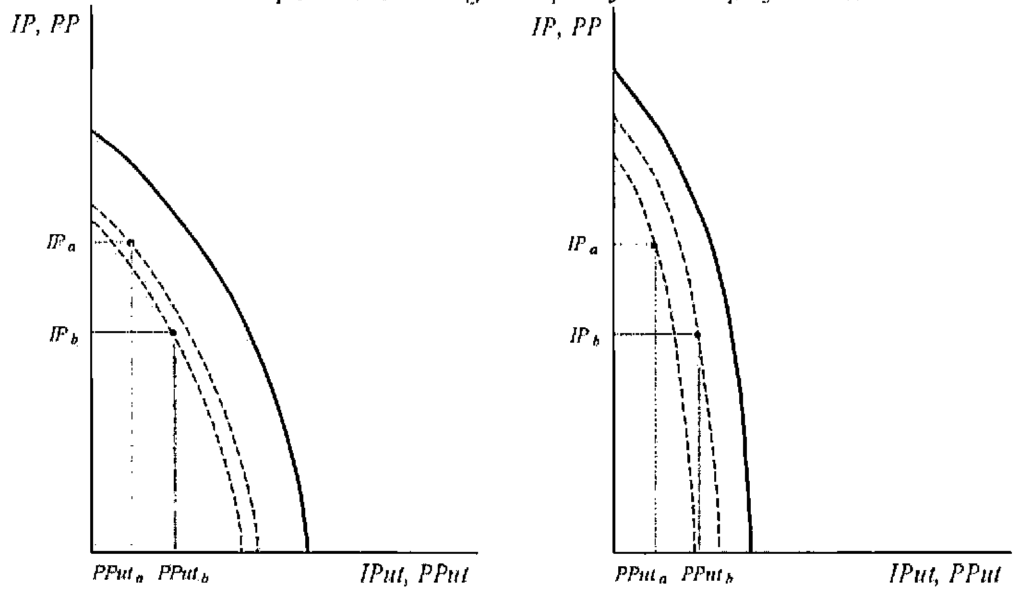


Figure IV.4

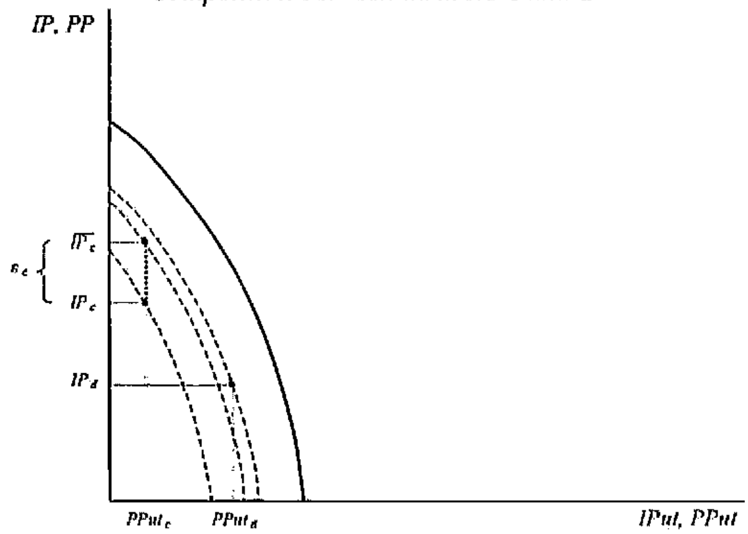
Insurers' competitive advantage and policyholder's preferences



Let's assume now that the insurance offer is granted by two different insurers (say *c* and *d*) who exhibit different equilibrium premiums, $IP_{p,c}^{*}$ and $IP_{p,d}^{*}$, so that insurer *c* enjoys a comparative advantage, as shown in Figure IV.5.

Figure IV.5

Competition between insurers c *and* d



Apparently, this insurance market is not different from the one depicted in Figure IV.3. However, a relevant difference concerning the possible reaction of the "disadvantaged" insurer (respectively, insurer *a* and *d*) exists. Assuming the insurers' cost structure is given and not changeable, while insurer *a* has only one alternative to compete with insurer *b* (*i.e.*, decreasing σ_a) insurer *d* has two alternatives to compete with insurer *c* (*i.e.*, both decreasing σ_d or increasing A_d/L_d).

The two ways of affecting the value of the insolvency put (changing σ_i or A_i/L_i) are quite different. Since σ_i depends on σ_A, σ_L (*i.e.*, the instantaneous standard deviations of assets and liabilities) and ρ (*i.e.*, the instantaneous correlation coefficient between assets and liabilities), a decrease of σ can be achieved by improving the underwriting activities or stabilising the investment income or through a closer matching between assets and liabilities.

On the contrary an increase of the *A/L* ratio is easier to obtain, since it requires just an equity increase.

Actually, the value of the insolvency put may be also affected by ceding the liabilities to a reinsurer. In such a case the reinsurer's risk of default should be taken into account to evaluate the policyholder's final risk exposition. In any case it is obvious that the lower the risk of default of the reinsurer, the lower the policyholder's final risk exposition.

Therefore this way of defining the insurer's strategies which affect risk seems consistent with Merton's classification of the ways to reduce default risk. According to Merton there exist three different ways for an intermediary with credit-sensitive activities to provide assurances against default risk to the customers who hold its liabilities[93]:

1) *hedging*: the firm holds assets which have payouts that "match" those promised on its contractual liabilities and it chooses a "transparent" structure so that customers can easily verify that such a matching policy is being followed;
2) *insuring*: the firm acquires guarantees of its customer liabilities from an AAA-credit-rated private-sector or government third party. The providing of such guarantees is a large financial-intermediation business, which is itself quite credit sensitive;
3) *building up capital cushions*: the firm raises additional capital beyond that required for the funding of the physical investments and working capital needed to run the intermediary.

For an insurance firm, the hedging strategy is someway related to the insurer's Asset-Liability Management (ALM) strategy, affecting the parameter ρ; the insuring strategy is related to the reinsurance program; and the building up capital cushions strategy is obviously related to the decisions concerning the insurer's leverage.

[93] See Merton (1997), p. 4.

IV.2.2. Opaqueness

The assumption underlying the previous arguments is that policyholders are able to evaluate the true value of the insolvency option.

As pointed out in Chapter II, provided that policyholders are averse towards insurers' riskiness, two conditions must hold so that policyholders may be able to evaluate insurers' riskiness, *i.e.*, in the framework suggested in Chapter III and IV, insurer's insolvency options: (1) there should be enough disclosure, and (2) costs of monitoring should be low. Neither condition can be assumed to be adequately satisfied.

As far as disclosure is concerned, many efforts have been done to improve transparency in the insurance sector.

The *Principles on Capital Adequacy and Solvency* paper, written out by the International Association of Insurance Supervisors, states that insurance companies should be required to disclose relevant information to the public[94].

> *In general, public disclosure of information regarding a company's solvency position enhances policyholders' and the market participants' ability to exercise discipline with respect to companies. Market discipline is enhanced by transparency of the solvency control level requirements and imposes strong incentives on insurers to conduct their business in a safe, sound and efficient manner. However, the sensitivity of the market to publicity regarding a company's solvency position should be considered when establishing disclosure requirements*[95].

As far as costs of monitoring are concerned, it is difficult to assume that policyholders are generally able to evaluate easily insurer's riskiness.

A note written out by the European Commission underlines that most of the policyholders

[94] See International Association of Insurance Supervisors (2002).

[95] Cf. International Association of Insurance Supervisors (2003), p. 8.

would not be able to benefit from public disclosures in order to "monitor" their insurance company. The following observation, which is taken from the banking document Basle Committee document from September 1998 "Enhancing bank transparency", can be applied to the insurance sector: "Also, public disclosure may have limited effectiveness in relation to banks with heavy reliance on retail deposits, since retail depositors may lack the training to monitor a bank's condition via its public disclosure"[96].

In a sense, policyholders can be assumed to be like bank depositors, facing opaque institutions and incurring too high cost of monitoring (compared to their stake in the bank).

As argued – among others – by Dewatripont and Tirole (1994), if depositors could effectively monitor their banks then this would limit socially suboptimal bank behaviour. But monitoring is expensive and there are limits to information. Moreover, bank debt is mainly hold by depositors who held only a small deposit and who lack the incentives to perform efficient monitoring. Thus, depositors need to be represented by a regulator who can intervene on their behalf to correct market failure.

Similarly, policyholders negotiating small value coverage may lack the incentives to perform efficient monitoring.

However, in many lines of business the selection of the insurance companies is not carried out directly by policyholders, but through the intervention of intermediaries, typically insurance brokers.

They should be regarded as much more able (because much more informed or much more skilled) than policyholders in evaluating the insurance firms' riskiness. Brokers could act in order to decrease both the effect of opaqueness and, under certain conditions, of the costs of monitoring. Assuming they act on behalf of the policyholders, brokers should exhibit similar preferences. In a sense, they could exhibit a greater aversion towards the insurers' insolvency, because they usually make substantial investments in relationships with their policyholder clients. As a consequence, the more involved professional advisors are in insurers' selection the more effective the market discipline mechanism should be.

[96] European Commission (2001), p. 11.

Moreover, since in many lines of business brokers are willing to negotiate only with insurers who have a certain minimum rating, the prevailing market discipline mechanism could work through quantity as the main signal. In other words, insurers with high risk of insolvency could not find any room in the market, whatever the price charged.
Similarly, when large risks are involved, it can be assumed that the market discipline mechanism can works more effectively, since policyholders have a greater stake in the insurer.
Finally, the costs of monitoring decreases (improving, in turn, the effectiveness of the market discipline mechanism) where insurers' ratings are easily available and prove to predict adequately insurers' insolvencies. In such a case, the rating would act as a proxy variable for the insolvency option. Actually, ratings are not very accurate predictors. As shown by Ambrose and Seward (1988), Ambrose and Carroll (1994) and Pottier (1998), the predictive ability of ratings can be improved if they are combined with other information, such as rating changes and financial ratios. Consequently the cost of solvency monitoring depends on both the quantity and quality of predictors variables used in insolvency prediction models and the way this information is used, *i.e.*, the party who performs the insolvency risk assessment.
In short, the market discipline mechanism should prove to be more efficient when:

- policyholders select insurers through skilled insurance advisers;
- risks to be insured are large;
- the predictive ability of ratings is high.

In any case, the hypothesis of perfect observability of the insolvency option is really far from describing a real insurance market. As with all the financial intermediaries, insurance firms exhibit little transparency towards their clients and investors. The opacity of insurance companies relates mainly to their liabilities. Their actual nature, value and riskiness are difficult to appraise by an external observer[97].

[97] See Ross (1989), p. 542;

If $PPut_{p,i}$ is not observable the previous representation of the policyholder's and insurer's behaviour does not hold anymore.

Policyholders will select policies according solely to the price $IP_{p,i}$, therefore they will look for coverages that exhibit the lowest price. Therefore, they could select, premiums being equal, the policy offered by the riskiest insurer, achieving a lower utility value then the one they could have achieved by negotiating, at the same price, with a less risky insurer.

Unobservability of $PPut_{p,i}$ not only affects policyholders behaviour, addressing them to the selection of the policy with a minimum $IP_{p,i}$, but could also get insurers themselves to became riskier in order to increase $IPut_{p,i}$ and decrease $IP_{p,i}$. In fact, the easiest way to do decrease $IP_{p,i}$ is through a reduction of A_i/L_i, which increases $IPut_{p,i}$.

The removal of the hypothesis of perfect observability of $PPut_{p,i}$ could bring about an incentive for insurers to assume default risk and therefore makes the aggregate level of policyholders' welfare worse. In addition, such an incentive could be stronger for less efficient insurers (i.e., insurers with high $Loading(E)_{p,i}$), which can ask adequately low prices only through a broad increase of their risk exposition (since they have to compensate $Loading(E)_{p,i}$).

While in a transparent context market forces succeeded in affecting insurance companies' behaviour, forcing them to reduce their risk exposition, information asymmetries can lead to a market failure, because the higher risk exposition of insurance companies, forced by the need of a price reduction, would transform policyholders (i.e., risk-avoiders into unaware risk-takers).

Opaque institutions are those for which monitoring, bonding, and control costs, i.e. agency problems, are most severe. Information is asymmetric between the participants and the institution, and the need for control and contractual structures to permit them to function is so critical that they essentially define these institutions. A depositor in an S&L has almost no knowledge of the particular loans the institutions is making. Are they local or national, secured or not? Little of this is visible to the participant. The same is true for the insurance company that invests premiums prior to paying claims (...).

IV.3. Implications for insurance supervision

IV.3.1. Perfect transparency

According to Cummins, in an economy where policyholders are assumed to have complete information about the default risk of competing insurers and to choose a company with a premium/default risk combination consistent with their tastes,

> *there is, of course, an interaction between the funds initially put up by stockholders and the value of the put;* i.e., *the put value is affected by firm capital structure*[98].

In this framework the optimal amount of an insurer's surplus depends on policyholders' preferences regarding insurer's riskiness.
Surplus becomes the easiest (but costly) tool that an insurer can use to maximise the probability to meet policyholder's preferences. Therefore, surplus acts as a risk-based capital, since its optimal amount depends on both the market preferences and the insurer's intrinsic risk. It could be said that it is the market discipline mechanism that correlates surplus to insurer's intrinsic risk.
According to Harrington (2004*b*), where a market discipline mechanism works effectively, stringent risk-based capital requirements could turn to be ineffective.
In fact, those requirements could affect the insurer's insolvency put in an inconsistent manner. Harrington (2004*a*) argues that

> *by requiring some firms to hold more capital in relation to their observed risk, risk-based capital rules may reduce their put values. With perfect information and costless enforcement, capital standards could correspond exactly with a firm's risk. Each firm could be forced to hold the efficient level of capital. But risk assessment is costly, inherently imperfect, and perhaps even relatively crude in the best of circumstances. Two types of errors are therefore inevitable with risk-based capital. First, capital standards will involve Type 1 errors; they will cause some otherwise adequately capitalized ("sound") firms (i.e., those with efficient put values) to hold*

[98] See Cummins (1988), pp. 829-830.

too much capital (or to reduce their risk inefficiently). Second, capital standards will involve Type 2 errors; they will fail to cause some firms with excessive put values to hold enough capital in relation to risk.

IV.3.2. Opaqueness

However, in a real-life economy, the hypothesis of perfect observability of insurers' riskiness does not hold. In such a case, both the arguments described in Section IV.2.2 and those suggested in other frameworks[99] show that insurers have an incentive to reduce the level of surplus. Therefore risk-based capital requirements become necessary in order to prevent market failures[100], even if they can give rise to the two types of errors described above.

In other words, capital standards may turn to be a second best solution in an perfect economy, but remain a first best solution in a real-life economy. Hence efforts should be made in order both to get the real-life economy close to a perfect one (through a strengthening of insurers' disclosure, forcing the system to produce timely and reliable information for use by management, supervisors, and market participants) and to improve the existing capital standards.

The goals of a properly designed risk-based capital system should be to provide incentives for insurers to reduce insolvency risk, to facilitate the rehabilitation of weak insurers, and to bring about the orderly exit of unsuccessful companies from the market. However, a poorly designed system could create incentives for excessive risk-taking by insurers, impose unnecessary costs on policyholders and financially sound insurers, and reduce market capacity.

Economic theory reveals that resource allocation and operating efficiency are best achieved in competitive markets. To attain these desirable economic objectives, solvency regulation should be designed to duplicate as closely as possible the

[99] For instance, Doherty (1989), pp. 269-270, shows that in a world where the policyholders cannot monitor the financial condition of the insurer (and hence the prices the policyholders are willing to pay for the policies offered by different companies are insensitive to their respective choices of leverage) the optimal level of surplus E, to be held for any given premium income P, is zero.

[100] Market failures would be induced by a competitive decrease of insurer's surplus, in order to reduce premiums. Beyond a certain extent insurers would become unable to meet they obligations.

outcome of a competitive and transparent market in which all parties have access to all relevant information. This means that solvency regulation should not attempt to prevent all insurers' failures. Market exit through failure and voluntary withdrawal is a normal outcome in a competitive market. However, failed firms should not be permitted to place a significant burden on policyholders or healthy insurers. Thus, solvency regulation should: (1) provide proper incentives for insurers to reduce insolvency risk, (2) facilitate, where possible, the rehabilitation of weak insurers, and (3) bring about the orderly exit of unsuccessful companies as closely as possible to the point where the economic value of assets falls below the economic value of liabilities. A risk-based capital requirements system is a form of solvency regulation and as such it should try to achieve these objectives.

In theory, risk-based capital refers to a system in which insurers would be required to meet or exceed a minimum capital requirement tied to specific characteristics of the companies that are presumably relates to the risk of insolvency, Under a risk-based capital system, if an insurer's reported capital (surplus) failed to exceed its required risk-based capital (or some percentage of risk-based capital), it would be subject to regulatory action. A well designed risk-based capital system should help regulators identify financially weak companies while there is still time for rehabilitation and remove unsalvageable companies from the market before they incur significant deficits that would harm policyholders and place a burden on healthy insurers. Such a system should also motivate insurers that otherwise would have inadequate incentives for satisfy to hold more capital and to reduce their risk of insolvency.

A formal risk-based capital system offers several possible advantages compared to the existing systems of minimum capital requirements and regulatory monitoring of capital relative to risk. These advantages have the potential to achieve an efficient reduction in the expected cost of insolvencies. As noted, a well-designed risk-based capital system would encourage greater safety for insurers for which market incentives for safety are inadequate. A well-designed risk-based capital system will also provide guidance and assistance to regulators. It will provide information to help identity weak companies and to facilitate regulatory intervention either before an insurer becomes insolvent or at a minimal level of deficit.

A risk-based capital system would give regulators legal authority to intervene if reported capital falls below risk-based capital requirements (or some percentage of risk-based capital). This authority will be valuable in cases where it might otherwise have been difficult for intervention to be upheld by the court system. In some instances, a risk-based capital system may force regulators to take some action rather than delay intervention due to pressure from the troubled company or the hope that things will get better without action.

On the other hand, risk-based capital requirements have a number of serious potential limitations. It is infeasible for a risk-based capital system to duplicate precisely the capital levels and incentives for safety that would exist in a competitive environment in which both consumers and insurers have adequate information and incentives for safety. Insolvency risk depends on numerous factors that are difficult to quantify, and the insurance market is characterised by substantial diversity across insurers in types of business written, characteristics of customers, and methods of operation. It is impossible to specify the "right" amount of capital for most insurers through a formula. Unavoidable imperfections in any meaningful risk-based capital system will likely distort some insurer decisions in undesirable and unintended ways. Overly stringent risk-based capital requirements would produce significant market dislocations to the detriment of many insurers and consumers. Thus, the desire to achieve the objectives of a risk-based capital system must be tempered by the reality that any such system will be imperfect and that the inevitable imperfections have the potential to impose significant costs on healthy insurers.

A poorly designed risk-based capital system could impose a variety of unnecessary costs on insurers and policyholders. Financially sound insurers operating in competitive markets provide insurance at the lowest possible costs, consistent with a given safety level, through diversification in their asset and liability portfolios and through the purchase of reinsurance. If the investment or underwriting risk charges imposed by a risk-based capital formula are not aligned with actual market risk levels, insurer investment and underwriting decisions could be distorted, leading to less effective diversification. This could actually reduce safety levels for financially sound insurers and lead to higher premiums rates for any given level of safety. A poorly designed system also could lead to unjustified damage to the reputations of well-managed insurers, raising the costs of capital for these firms and impeding their ability to raise new equity capital. The result could be a reduction in industry capacity and a shift toward self-insurance and other risk management alternatives. To the extent that risk management could be provided more effectively by a competitive insurance market, overall risk management costs would rise and risk management would become less effective. In addition, risk-based capital requirements by themselves will do little or nothing to help regulators determine whether an insurer's reported net worth is overstated. The great difficulty on determining whether an insurer's reported losses and losses reserves are significantly understated, especially for long-tailed lines with highly volatile costs, limits the ability of risk-based capital to encourage weak insurers to hold more capital and to assist regulators. In fact, poorly designed risk-based capital requirements could increase incentives for some insurers to under-report loss reserves in order to show lower required risk-based capital, higher capital relative to required risk-based capital, or both. Some insurers will try to manage their required level of risk-based capital through means that do not reduce risk or increase economic net worth. Thus, risk-based capital is not a substitute for regulatory monitoring of prices, reserves, and other financial variables.

References

Ambrose J. M., Carroll A. M. (1994), "Using Best's Ratings in Life Insurer Insolvency Prediction", in *Journal of Risk and Insurance*, Vol. 61, pp. 317-327.

Ambrose J. M., Seward A. (1988), "Best's Ratings, Financial Ratios and Prior Probabilities in Insolvency Prediction", in *Journal of Risk and Insurance*, Vol. 55, pp. 229-244.

Arrow K.J. (1965), *Aspects of the Theory of Risk-Bearing*. Helsinki, Yrjö Hahnsson Foundation.

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

Babbel D. F. (1998), *Components of Insurance Firm Value and the Present Value of Liabilities*, Working Paper, 98-18, The Wharton School, University of Pennsylvania.

Babbel D. F., Gold J., Merrill C. B. (2002), "Fair Value of Liabilities: The Financial Economics Perspective", in *North American Actuarial Journal*, Vol. 6, pp. 1-11.

Beard R. E., Pentikäinen T., Pesonen E. (1984), Risk Theory, The Stochastic Basis of Insurance, 3rd edition, London, Chapman & Hall.

Beltratti A. (2000), I mercati finanziari, Roma, Carocci.

Berger L. A., Cummins J. D., Tennyson S. (1992), Reinsurance and the Liability Insurance Crisis, in *Journal of Risk and Uncertainty*, Vol. 5, pp. 253-272.

Billett M., Garfinkel J., O'Neal E. (1998), "The Cost of Market versus Regulatory Discipline in Banking", in *Journal of Financial Economics*, Vol. 48, pp. 333-358.

Bliss R. R. (2001), "Market Discipline and Subordinated Debt: A review of Some Salient Issues", in Mishkin F. S. (editor), *Prudential Supervision: What Works and What Doesn't*, Federal Reserve Bank of Chicago Economic Review, pp. 24-45.

Bliss R. R., Flannery M. J. (2001), "Market Discipline in Governance of U.S. Bank Holding Companies: Monitoring vs. Influencing", in Mishkin F. S. (editor), *Prudential Supervision: What Works and What Doesn't*, Chicago, IL, The University of Chicago Press

Blum J. (2002), "Subordinated Debt, Market Discipline, and Banks' Risk Taking", in *Journal of Banking and Finance*, Vol. 26, pp. 1427–1441.

Board of Governors of the Federal Reserve System (1999), *Using Subordinated Debt as an Instrument of Market Discipline*, Report of a Study Group on Subordinated Notes and Debentures, Staff Study, No. 172.

Borch K. H. (1974), *The Mathematical Theory of Insurance*, Lexington, Lexington Books.

Brewer E., Mondschean T., Strahan P. (1997), "The Role of Monitoring in Reducing the Moral Hazard Problem Associated with Government Guarantees: Evidence from the Life Insurance Industry", in *Journal of Risk and Insurance*, Vol. 64, pp. 302-322.

Bühlmann H. (1970), *Mathematical Methods in Risk Theory*, Heidelberg, Springer.

Butsic R. P. (1994), "Solvency Measurement for Property-Liability Risk-Based Capital Applications", in *Journal of Risk and Insurance*, Vol. 61, pp. 656-690.

Cook P. J., Graham D. A. (1977), "The Demand for Insurance and Protection: The Case of Irreplacable Commdities", in *Quarterly Journal of Economics*, Vol. 91, pp. 143-156.

Cordella T., Yeyati E. L. (1998), *Public Disclosure and Bank Failures*, CEPR Discussion Paper No. 1886.

Crabbe L., Post M. (1994), "The Effect of a Rating Downgrade on Outstanding Commercial Paper" in *Journal of Finance*, Vol. 49, pp. 39-56.

Cummins J. D. (1988), "Risk-Based Premiums for Insurance Guaranty Funds", in *Journal of Finance*, Vol. 43, pp. 823-839.

Cummins J. D. (1992), "Financial Pricing Models of Property and Liability Insurance", in Dionne G. (Ed.), *Contributions to Insurance Economics*, Norwell, MA, Kluwer Academic Publishers.

Cummins J. D. (2000), "Allocation of Capital in the Insurance Industry", in *Risk Management and Insurance Review*, Vol. 3, pp. 7-28.

Cummins J. D., Danzon P. M. (1997), "Price, Financial Quality, and Capital Flows in Insurance Markets", in *Journal of Financial Intermediation*, Vol. 6, pp. 3-38.

Cummins J. D., Doherty N. A., Lo A. (2002), "Can Insurers Pay for the "Big One"? Measuring the Capacity of the Insurance Market to Respond to Catastrophic Losses", in *Journal of banking and Finance*, Vol. 26, pp. 557-583.

Cummins J. D., Mahul O. (2003), "Optimal Insurance With Divergent Beliefs About Insurer Total Default Risk", in *Journal of Risk and Uncertainty*, Vol. 27, pp. 121-138.

Cummins J. D., Nini Gregory P. (2002), "Optimal Capital Utilization by Financial Firms: Evidence from the Property-Liability Insurance Industry", in *Journal of Financial Services Research*, Vol. 21, pp. 15-53.

Cummins J. D., Phillips R. D. (2000), "Applications of Financial Pricing Models In Property-Liability Insurance", in Dionne G. (ed.), *The Handbook of Insurance Economics*, Boston, Kluwer Academic Publishers.

Cummins J. D., Sommer D. W. (1996), "Capital and Risk in Property-Liability Insurance Markets", in *Journal of Banking & Finance*, Vol. 20, No. 6, July, pp. 1069-1092.

Danthine J.-P., Donaldson J. B. (2002), *Intermediate Financial Theory*, Upper Saddle River, NJ, Prentice Hall.

Demsetz R. S., Saidenberg M. R., Strahan P. E. (1996), "Banks with Something to Lose: the Disciplinary Role of Franchise Value", in *Economic Policy Review*, Vol. 2, pp. 1-14.

Dewatripont M., Tirole J. (1994), *The Prudential Regulation of Banks*, MIT Press.

Diamond D., Dybvig P. (1983), "Bank Runs, Deposit Insurance and Liquidity", in *Journal of Political Economy*, Vol. 91, pp. 401–19.

Doherty N. (1989), "On the Capital Structure of Insurance Firms", in Cummins J. D., Derrig R. A. (eds.), *Financial Models of Insurance Solvency*, Norwell, MA, Kluwer Academic Publishers.

Doherty N. A., Garven J. R. (1986), "Price Regulation in Property-Liability Insurance: A Contingent-Claims Approach", in *Journal of Finance*, Vol. 41, pp. 1031-1050.

Doherty N., Schlesinger H.(1990), "Rational Insurance Purchasing: Consideration of Contract Non-Performance ", in *Quarterly Journal of Economics*, Vol. 105, pp. 243-253.

Doherty N. Tinic S. (1981), "Reinsurance Under Conditions of Capital Market Equilibrium: A Note", in *Journal of Finance*, Vol. 36, pp. 949-953.

Downs D. H., Sommer D. W. (1999), "Monitoring, Ownership, and Risk-Taking: The Impact of Guaranty Funds", in *Journal of Risk and Insurance*, Vol. 66, pp. 477-497.

Elton E. J., Grubr M. J. (1995), *Modern Portfolio Theory and Investment Analysis*, 5th edition, New York, John Wily & Sons, Inc.

Epemarnis K., Harrington S.E. (2004), "Market Discipline in Property/Casualty Insurance: Evidence from Premium Growth Surrounding Policyholder Rating Changes", forthcoming.

European Commission (2001), *Banking rules: relevance for the insurance sector?*, DG Internal Market, Financial Institutions – Insurance, Note to the Solvency Subcommittee, downloadable at http://europa.eu.int/comm/internal_market/insurance/docs/markt-2056/markt-2056-01_en.pdf.

Fairley W. B. (1979), "Investment Income and Profit Margins in Property-Liability Insurance: Theory and Empirical Results", in *Bell Journal of Economics*, Vol. 10, pp. 192-210.

Fenn G. W., Cole R. A. (1994), "Announcements of Asset-Quality Problems and Contagion Effects in the Life Insurance Industry", in *Journal of Financial Economics*, Vol. 35, pp. 181-198.

Flannery M. (1998), "Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence", in *Journal of Money, Credit and Banking*, Vol. 30, pp. 273-305.

Flannery M. J. (2001), "The Faces of "Market Discipline"", in *Journal of Financial Services Research*, Vol. 20, pp. 107-119.

Flannery M., Sorescu S. (1996), "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983-1991", in *Journal of Finance*, Vol. 51, pp. 1347-1377.

Friedman M., Savage L. J. (1948), "The Utility Analysis of Choices Involving Risk", in *Journal of Political Economy*, Vol. 56, 279-304.

Froot K.A. (2003), *Risk Management, Capital Budgeting and Capital Structure Policy for Insurers and Reinsurers*, Preliminary version of the paper presented at the Geneva Risk Lecture at the 30th Seminar of the European Group of Risk and Insurance Economists, Zurich.

Gerber H. U., Pafumi G. (1998), "Utility Functions: From Risk Theory to Finance", in *North American Actuarial Journal*, Vol. 2, No. 3, pp. 74-90.

Girard L. N. (2002), "An Approach to Fair Valuation of Insurance Liabilities Using the Firm's Cost of Capital", in *North American Actuarial Journal*, Vol. 6, No. 2, pp. 18-46.

Goldberg, L., Hudgins S. (2002), "Depositor Discipline and Changing Strategies for Regulating Thrift Institutions", in *Journal of Financial Economics*, Vol. 63, pp. 263-274.

Harrington S.E. (2004*a*), "Capital Adequacy in Insurance and Reinsurance", forthcoming.

Harrington S.E. (2004*b*), "Market Discipline in Insurance and Reinsurance", forthcoming.

Harrington S. E., Danzon P. M. (1994), "Price Cutting in Property-Liability Insurance", in *Journal of Business*, Vol. 67, pp. 511-538.

Herring R. J., Vankudre P. (1987), "Growth Opportunities and Risk Taking by Financial Intermediaries", in *Journal of Finance*, Vol. 42, pp. 583-599.

Hogg R. V., Klugman S. A. (1984), *Loss Distribution*, New York, John Wily & Sons, Inc.

Hill R. D. (1979), "Profit Regulation in Property-Liability Insurance", in *Bell Journal of Economics*, Vol. 10, pp. 179-191.

International Association of Insurance Supervisors (2002), *Principles on Capital Adequacy and Solvency*, Principles No. 5, downloadable at http://www.iaisweb.org/02solvency.pdf.

International Association of Insurance Supervisors (2003), *Solvency Control Levels Guidance Paper*, Guidance Paper No. 6, downloadable at http://www.iaisweb.org/185solvency03.pdf.

Johnson H., Stulz R. (1987), "The Pricing of Options with Default Risk", in *Journal of Finance*, Vol. 42, pp. 2267-2280.

Kahane Y., Tapiero C. S., Laurent J. (1989), "Concepts and trends in the study of insurer's solvency", in Cummins J. D., Derrig R. A. (eds), *Financial Models of Insurance Solvency*, Norwell, MA, Kluwer Academic Publishers.

Kahneman D., Tversky A. (1979), "A Prospect Theory: An Analysis of Decision under Risk", in *Econometrica*, Vol. 47, pp. 263-291.

Keeley M. (1990), "Deposit Insurance, Risk, and Market Power", in American Economic Review, Vol. 80, pp. 1183-1200.

Kraus A., Ross S. A. (1982), "The Determination of Fair Profits for the Property-Liability Insurance Firm", in *Journal of Finance*, Vol. 37, 1015-1028.

Kwan S. (2002), "Bank Security Prices and Market Discipline", *Federal Reserve Bank of San Francisco Economic Letter*, No. 37, December.

Lane T. D. (1993), "Market Discipline", *IMF Staff Paper*, N. 40, pp. 53-88.

Lee S.-J., Mayers D., Smith C. W. (1997), "Guaranty Funds and Risk-Taking: Evidence from the Insurance Industry", in *Journal of Financial Economics*, Vol. 44, pp. 3-24.

Lee S.-J., Smith M. (1999), "Property-Casualty Insurance Guaranty Funds and Insurer Vulnerability to Misfortune", in *Journal of Banking and Finance*, Vol. 23, pp. 1437-1456.

Llewellyn D. T., Mayes D. G. (2003), *The role of market discipline in handling problem banks*, Bank of Finland Discussion Papers, No. 21.

Llewellyn D. T. (2002), "Alternative Approaches to Regulation and Corporate Governance in Financial Firms", in Johnson O. E. G. (editor), *Financial Risks, Stability and Globalisation*, Washington, DC, International Monetary Fund.

Machina M. (1987), "Choice Under Uncertainty: Problems Solved and Unsolved", in *Journal of Economic Perspectives*, Vol. 1, pp. 121-154.

Mayers D., Smith C. W. (1983), "The Interdependence of Individual Portfolio Decisions and the Demand for Insurance", in *Journal of Political Economy*, Vol. 91, pp. 304-311.

Merton R. C. (1989), "On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance", in *The Geneva Papers on Risk and Insurance*, Vol. 14, pp. 225-262.

Merton R. C. (1995), "A Functional Perspective of Financial Intermediation", in *Financial Management*, Vol. 24, pp. 23-51.

Merton R. C. (1997), "A Model of Contract Guarantees for Credit-Sensitive, Opaque Financial Intermediaries", in *European Finance Review*, vol. 1, pp. 1–13.

Merton R.C., Perold A.F. (1993), "Theory of Risk Capital in Financial Firms", in *Journal of Applied Corporate Finance*, Fall, pp.16-30.

Moody's (1998), *Property and Casualty Insurance and Reinsurance*, New York, Moody's Investors Service.

Mossin J. (1968), "Aspects of Rational Insurance Purchasing", in *Journal of Political Economy*, Vol. 76, pp. 553-568.

Myers S. C., Cohn R. A. (1987), "A discounted Cash Flow Approach to Property-Liability Insurance Rate Regulation" in Cummins J. D., Harrington (eds.), *Fair Rate of Return in Property-Liability Insurance*, Norwell, MA, Kluwer Academic Publishers.

Phillips R. D., Cummins J. D., Allen F. (1998), "Financial Pricing of Insurance in the Multiple Line Insurance Company", in *Journal of Risk and Insurance*, Vol. 95, pp. 597-636.

Pottier S. W. (1998), "Life Insurer Financial Distress, Best's Ratings and Financial Ratios", in *Journal of Risk and Insurance*, Vol. 65, pp. 275-288.

Pratt J. W. (1964), "Risk Aversion in the Small and in the Large", in *Econometrica*, Vol. 32, pp. 122-136.

Ross S. A. (1989), "Institutional Markets, Financial Marketing, and Financial Innovation", in *The Journal of Finance*, Vol. 44, pp. 541-556.

Schlesinger H. (2000), "The Theory of Insurance Demand", in Dionne G. (ed.), *The Handbook of Insurance Economics*, Boston, Kluwer Academic Publishers.

Schlesinger H., Doherty N. (1985), "Incomplete Markets for Insurance: An Overview", in *Journal of Risk and Insurance*, Vol. 52, pp. 402-423.

Singh A. K., Power M. L. (1992), "The Effects of Best's Rating Changes on Insurance Company Stock Prices", in *Journal of Risk and Insurance*, Vol. 59, pp. 310-317.

Sironi A. (2001), "An Analysis of European Banks' SND Issues and its Implications for the Design of a Mandatory Subordinated Debt Policy", in *Journal of Financial Services Research*, Vol. 20, pp. 233-266.

Sironi A. (2003), "Testing for Market Discipline in the European Banking industry: Evidence from Subordinated Debt Issues", in *Journal of Money, Credit and Banking*, Vol. 35, pp. 443-472.

Smith C. W. (1979), "Application of Option Pricing Analysis", in Bicksler J. L. (ed.), *Handbook of Financial Economics*, Amsterdam, North-Holland Publishing Company.

Sommer D. (1996), "The Impact of Firm Risk on Property-Liability Insurance Prices", in *Journal of Risk and Insurance*, Vol. 63, pp. 501-514.

Tapiero C., Kahane Y., Jacque L. (1986), "Insurance Premiums and Default Risk in Mutual Insurance", in *Scandinavian Actuarial Journal*, pp. 82-97.

von Neumann J., Morgenstern O. (1953), *Theory of Games and Economic Behavior*, 3rd edition, Princeton, NJ, Princeton University Press.

Zanjani G. (2002*a*), "Pricing and capital allocation in catastrophe insurance", in *Journal of Financial Economics*, Vol. 65, pp. 283-305.

Zanjani G. (2002*b*), *Market Discipline and Government Guarantees in Life Insurance*, Draft, Federal Reserve Bank of New York.

APPENDIX

The option pricing theory as a framework to evaluate the insurance policies: the insolvency option

In determining the price of insurance, traditional actuarial models assume that premiums are primarily determined by the insurer, according to the so-called "supply-side perspective". The drawback of this approach is that it fails to take into account the role played by markets in determining the price of insurance. Starting in the 1970s, financial methods of pricing insurance have begun to try to fill this gap, investigating the insurance valuation problem in a financial market context, assuming that premiums should reflect equilibrium relationships between risk and return or avoid the creation of arbitrage opportunities.

According to financial theory, insurance contracts can be viewed as roughly analogous to the bonds issued by non-financial corporations[101]. Therefore insurance financial pricing models have been developed to price this special class of liabilities using various specific frameworks, such as the capital asset pricing model[102], the arbitrage pricing theory[103], the discounted cash flow model[104]. In all these frameworks the insurer's default risk is not explicitly or implicitly incorporated into the valuation model. However, the idea that the evaluation of an insurance policy should also depend upon the insurer's solvency is quite

[101] Even if, as pointed out by Cummins (1992), p. 142, insurance liabilities are not like conventional bonds but more like structured securities, where payoffs are triggered by various contingencies. The payment times and amounts for insurance policies are stochastic, determined by contingent events.

[102] Cf. Fairley (1979) and Hill (1979).

[103] Cf. Kraus, Ross (1982).

[104] Cf. Myers, Cohn (1987).

reasonable and in fact the model of Doherty and Garven (1986) and Cummins (1988), which incorporates default risk to value insurance by using option pricing theory, has tried to remedy this deficiency in the literature[105]. From then on many works, prompted by different purposes, have made use of the option pricing theory[106].

The standard option pricing model of insurance views the liabilities created by issuing insurance policies as analogous to risky corporate debt. The insurer is assumed to issue an insurance policy in return for a premium payment, analogous to the proceeds of a bond issue. In return, it promises to make a claim payment to the policyholders at the maturity date of the contract.

The simplest option interpretation of the firm involves a one-period, two-date model where the firm issues policies at time 0 and claim payments occur at time 1.

The insurer is assumed to begin operations at time 0 with a surplus of E. It issues insurance policies (liabilities) with a value of L. Assets ($A = L + E$) and liabilities are assumed to follow geometric diffusion processes:

$$\frac{dA}{A} = \alpha_A dt + \sigma_A dz_A$$

$$\frac{dL}{L} = \alpha_L dt + \sigma_L dz_L$$

[105] Actually, an early application of the option pricing theory can be found in Smith (1979), pp. 106-108, who proposes a Black-Scholes put option model for insurance pricing. The proposed contract offers insurance against the decline in the value (V) of an asset below the insured value (X). The insurance premium is given by the formula:

$$P = -VN\left[\frac{-\ln(V/X)-\left(r+\sigma^2/2\right)\tau}{\sigma\sqrt{\tau}}\right] + Xe^{-r\tau}N\left[\frac{-\ln(V/X)-\left(r-\sigma^2/2\right)\tau}{\sigma\sqrt{\tau}}\right]$$

where τ, σ^2 are the time to expiration and risk parameter of the insured asset and $N[\cdot]$ is the standard normal distribution function. Smith suggests that this model might be applicable to mortgage insurance.

[106] An incomplete list of the references related to the option pricing theory in the evaluation of the insurance contracts includes the following works: Doherty (1989), p. 269; Butsic (1994), p. 676; Phillips, Cummins, Allen (1998), pp. 600-607; Cummins, Sommer (1996), pp. 1072-1073; Sommer (1996), pp. 502-504; Cummins, Danzon (1997), p. 5; Cummins (1992), pp. 150-161; Babbel (1998), pp. 2-4; Downs, Sommer (1999), p. 479; Cummins (2000) pp. 8-13; Girard (2002),

where $dz_L dz_A = \rho dt$, α_A, α_L are the instantaneous expected changes in assets and liabilities, respectively; σ_A and σ_L are the instantaneous standard deviations of assets and liabilities, and ρ is the instantaneous correlation coefficient between assets and liabilities.

If assets exceed liabilities at time 1 (*i.e.*, $A_1 > L_1$), the firm pays the losses and the equity holders receive the residual value (the difference between assets and liabilities, *i.e.*, $E_1 = A_1 - L_1$). However, if assets are less than liabilities, the insurer defaults and the policyholders receive the assets. The payoff to policyholders at time 1 is thus equal to $\min[A_1, L_1]$, which in turn is equal to $L_1 - \max[L_1 - A_1, 0]$. Since $\max[L_1 - A_1, 0]$ is the payoff of a put option, the present value of the policyholders' claim prior to the liability can be expressed as the riskless present value of the liabilities[107] minus the premium of this put, that is:

$$\text{Value of policyholders' claim} = Le^{-r\tau} - Put(A, L, r, \tau, \sigma) \qquad \text{(A.1)}$$

where $Put(A, L, r, \tau, \sigma)$ is the value of a put option on A, evaluated at time interval $\tau \in [0,1]$ from the maturity, with strike price L, interest rate r (equal to the risk-free rate, r_f, minus the inflation rate for liabilities, r_L[108]) and risk parameter $\sigma^2 = \sigma_A^2 + \sigma_L^2 - 2\rho\sigma_A\sigma_L$[109].

The put represents a discount in the price of the insurance which reflects the expected value of the equity holders' option to default if $A < L$ and is the *insolvency put*. Thus, the fair price of the insurance is the riskless present value of losses less the insolvency put.

The insolvency put option value arises from the limited liability enjoyed by equity holders when the firm issues debt (*i.e.*, insurance policies, the major debt of an insurance company). It is the value to equity holders of capturing the upside

---

p. 23; Cummins, Nini (2002), p. 22; Babbel, Gold, Merrill (2002), pp. 14-16; Cummins, Phillips (2000), pp. 633-645.

[107] *I.e.*, the present value of the liabilities if default risk were zero.

[108] The meaning of r_L and its link with α_L is clearly defined in Phillips, Cummins, Allen (1998), p. 601.

earnings while not incurring all the downside costs of default. The insolvency put option increases in value as the insurer takes on more risk. If the insurer faces minimal insolvency risk, he may take little advantage from this component of value; but if it is a risky firm, the implicit insolvency put option may be of considerable value.

There are two endogenous variables that affect the value of the insolvency put, the asset-to-liability ratio A/L and the risk parameter σ. The insolvency put value varies inversely with the asset-to-liability ratio and directly with the risk parameter. By choosing appropriate values for these two parameters, the insurer can achieve its target level of expected insolvency costs. According to this approach, the greater the insolvency risk, the greater the value of the insolvency put and thus the lower the policyholder's value of the insurance. If insurance prices reflected the true value of the insurance purchased, then riskier firms should receive lower premiums for their insurance than safer firms. One of the aim of the following sections is to show how this result can be explained within the framework of a microeconomic analysis of the policyholders' and insurers' behaviour.

One of the direct applications of option modelling in insurance is a deeper analysis of the insolvency risk. This application utilises the put-call parity formula:

$$A = Call(A, L, r, \tau, \sigma) + [Le^{-r\tau} - Put(A, L, r, \tau, \sigma)] \quad \text{(A.2)}$$

where $Call(A, L, r, \tau, \sigma)$ and $Put(A, L, r, \tau, \sigma)$ are, respectively, the values of a call and a put option on A, evaluated at time interval $\tau \in [0,1]$ from the maturity, with strike price L, interest rate r and risk parameter σ^2. Since A = Value of policyholders' claim + Value of equity holders' claim, (A.1) and (A.2) give

[109] Cummins (1988), cf. p. 829, demonstrates that, for property-liability insurers, the correct risk parameter includes components for asset risk, liability risk, and the correlation among assets and liabilities.

not reflected in its contract prices. However these conclusions may change if it is assumed that a sort of financial monitoring exists so that insurance premiums impound some information about the insurer's risk of default.