Risk transfer and agency risk: some examples of mathematical modelling in law & economics*

Elisa Luciano
International Center for Economic Research, Turin,
University of Turin, Dept. Statistics & Mathematics
& FERC, Warwick University, UK

July 2004

There are efforts in diverse fields of law [..] to deal with facts on a more systematic and objective basis: It is natural that these efforts to make law more "scientific" should involve mathematics [..] Most attempts to use quantitative methods in litigated disputes have encountered strong legal and technical objections. Clearly the effort is not worth making unless mathematical findings are a significant aid to legal judgement: this is indeed the case in diverse fileds of law [..] The resolving power of a mathematical approach is very high, and it brings into focus our inchoate ideas and purposes, just as we gain perspectives of a different sort from a study of foreign codes or historical processes.

M. O. Finkelstein, 1978

This paper explores several applications of mathematical methods to law-and-economics problems, focusing on two aspects of legal contracts: the transfer of risk underlying the vast majority of contracts and agency risk.

It begins with the observation that firms frequently face excessive risks, especially when the latter are computed according to current international standards. Two types of contracts aimed at mitigating risks are therefore analyzed, namely insurance and financial (derivative) ones. After having outlined the features of these contracts, we proceed to discuss a typical problem of contract fulfilment: the so-called principal agent relationship. The latter arises in any situation where one party to the contract (the agent) hired to perform some task can affect the outcome for the counterpart (the principal).

The paper is organized as follows: Section 1 discusses risk measurement and management issues in general, taking for granted that risk transfer is what

*Material prepared for the Doctorate in Economics of Law & Institutions, Cornell, Gent, Turin Universities and Ecole Polytechnique, Paris. The author is indebted with Lorenzo Peccati for extremely helpful suggestions in preparing this material, for being allowed to use his teaching material and, above all, for his brilliant example and encouragment in teaching maths to students in law. Marco Taboga provided assistantship on an eralier version of the paper.

underlies most contracts. It formalizes the most common (non-industrial) risk measure, i.e. Value-at-Risk. The subsequent sections deal with risk transfer tools. In particular, section 2 establishes the rationale of insurance in risk mitigation, while introducing some insurance contracts and their coverage on the part of the insurance company. Section 3 covers risk management via financial rather than insurance contracts: it deals with derivative securities and their valuation in efficient markets. Section 4 introduces the other main issue we are concerned with: the agency problem in contract fulfilment, and the ensuing design of incentive schemes. Section 5 explores some agency issues extracted from recent law & economics literature.

Our presentation does not aim to be complete, but is targeted at convincing the reader that a correct formalization of factual problems can help disentangle not only their solution, but also the main drivers. Knowledge of basic calculus, notions of probability and random variable (rv) is required.

Most of the applications covered in sections 1 to 3 are drawn from Peccati (2002). The reader is also referred to Filkenstein (1978) for a comprehensive coverage of mathematical formalization of judicial problems.

1 Risk issues and measurement

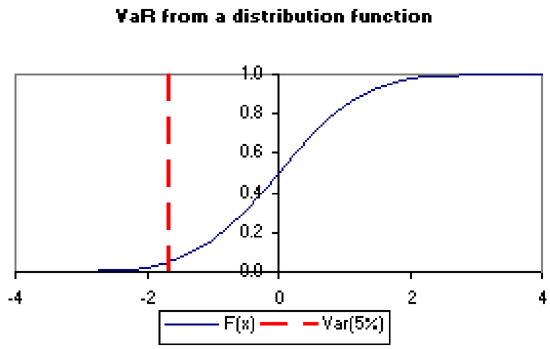
The underlying object of most contemporary legal contracts is economic risk. We can divide economic risks into two categories: industrial and financial, with the latter including, for instance:

- market risk
- credit risk
- operational (including legal) risk[1]
- default risk

The measurement of these risks is perceived to be a crucial issue, both in order to establish the amount of capital coverage (safety capital) and to eventually decide reduction actions (risk transfer, the so called risk management). Safety capital in turn can be either internally required or comply with the regulation (capital requirement).

It is therefore appropriate to begin our brief survey of mathematical applications to legal problems with some issues of risk measurement and management. In particular, we focus on measurement via Value-at-Risk and on management via the purchase of insurance and financial contracts.

[1] Operational risk in this context means risk of inadequacy in information or risk measurement systems for organizational or technical reasons, including fraud, technology risks, risk of lawsuits by shareholders, compliance, and regulatory risk.



Supranational and national authorities worldwide, with the BIS as a leader, have progressively adopted a measure called Value at Risk (VaR). VaR is the maximum loss which can occur, at a given level of confidence and over a given time horizon.[2]. It is used as a standard to control the risks on financial assets and credit exposures (market and credit risk respectively), as well as operational risks. In the interest of clarity, we will consider it in this context of market risk, i.e. referring to the potential losses on a financial asset or portfolio of assets, due to changes in prices.

Formally, let us denote the (absolute or relative) change in the portfolio value at a given horizon T as a random variable $\mathbf{X}$ and its distribution function as $F(x) = \Pr(\mathbf{X} \leq x)$. Let us fix also a confidence level β, and denote its complement to one as α.We are interested in finding the threshold (real number) VaR such that the gain or loss $\mathbf{X}$ will be lower than VaR with probability α.

$$\Pr(\mathbf{X} \leq VaR) = F(VaR) = 1 - \beta = \alpha$$

For instance

$$\Pr(\mathbf{X} \leq VaR_{.95}) = 5\%$$

With F strictly increasing, endowed with a density function $f, F(VaR) = \alpha$ means

$$\int_{-\infty}^{VaR} f(x)dx = \alpha$$

VaR has been introduced as a risk indicator in contractual provisions and financial management by the JP Morgan's CEO, Weatherstone, in the early nineties. Subsequently, there have been two main incentives for its world-wide adoption:

1. financial derivatives debacles

[2]A comprehensive survey of VaR is Jorion (1996). The regulatory interventions can be found at the BIS web page, www.bis.org.

2. lobbies and international regulations (BIS, ISDA, G30)

Among the former, we usually recall the disasters of

- Orange County, USA, amounting to1640 $ million, and due to borrowing short and lending long;
- Barings, UK, amounting to 1330 $ million and due to lack of coordination between front desk and back office in stock index futures;
- Metallgesellschaft, D, arriving at1340 $ million, due to a rolling hedge using 3-months futures against 10-year forwards in the oil market.

In spite of its recent "discovery" by management, VaR is a very well known statistical measure, namely the $\alpha-$quantile of $\mathbf{X}$.

If we knew the analytical expression for F, calculating VaR would be extremely easy, since from the definition, $F(VaR) = \alpha$, we would immediately obtain, by inversion (or generalized inversion), $VaR = F^{-1}(\alpha)$. However, in daily usage the analytics for the distribution of returns (profits and losses) is unknown: one needs therefore an assumption on F, such as "F is normal", followed by an estimate of its parameters[3]. Techniques of best fitting can be adopted in order to select the "best" F among a finite group: however, the normality assumption is the most common one.

If $\mathbf{X}$ is normally distributed, with expected value m and variance σ^2, we know that $\mathbf{Y} = (\mathbf{X} - m)/\sigma$ has a standard normal distribution, i.e.,

$$\Pr(\mathbf{Y} \leq y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z \tag{1}$$

Let us denote this distribution function as $\Phi(y) = \Pr(\mathbf{Y} \leq y)$.

From the VaR definition, $\Pr(\mathbf{X} \leq VaR) = \alpha$, we get:

$$\Pr\left(\frac{\mathbf{X} - m}{\sigma} \leq \frac{VaR - m}{\sigma}\right) = \alpha$$

or:

$$\Pr\left(\mathbf{Y} \leq \frac{v - m}{\sigma}\right) = \Phi\left(\frac{v - m}{\sigma}\right) = \alpha$$

and finally:

$$v = m + \sigma\Phi^{-1}(\alpha) = m + \sigma\Phi^{-1}(1 - \beta)$$

For any given level of confidence β, $\Phi^{-1}(1-\beta)$ is reported in statistical tables: for instance, when $\beta = 99\%, \Phi^{-1}(1-\beta) = -2.33$, while, when $\beta = 95\%, \Phi^{-1}(1-\beta) = -1.6449$.

Usually, financial authorities fix a level of confidence β, and therefore a value of $\Phi^{-1}(1-\beta)$: financial institutions as well as non financial firms must

[3] Alternative techniques for VaR estimation (Monte Carlo and hystorical simulation) will not be considered here: see for instance Jorion (1996).

determine only the mean and standard deviation of their returns, m and σ, in order to calculate their risk exposure.

The normality assumption on returns over a specific horizon, say one year, can imply that returns over subhorizons, such as one month, are still normal, with mean equal to $m/12$ and standard deviation $\sigma/\sqrt{12}$. This, together with the above note on β, allows us to easily adapt VaR measurement to any level of confidence and any horizon, starting only from an estimate of mean returns and their volatility over a fixed horizon.

Example — Suppose profit/losses over one year $\mathbf{X}$ are normal, with expected value $m = 15$ million \$ and variance $\sigma^2 = 100$ millions. We want to calculate the VaR with a confidence level of 95%. The standard deviation is $\sigma = \sqrt{100} = 10$ millions. The VaR is:

$$VaR_{95\%} = m - 1.644\,9\sigma = 15 - 1.644\,9 \times 10 = -1.449 \text{ millions}$$

That is, in 95% of the possible cases the portfolio will not lose more than 6.449 millions.

If we increase the level of confidence to $\beta = 99\%$, the VaR becomes:

$$VaR_{99\%} = m - 2.326\,3\sigma = 15 - 2.326\,3 \times 10 = -8.326 \text{ millions}$$

By increasing the level of confidence, we have obtained a more prudential valuation of the maximum portfolio loss.

By maintaining the level of confidence to 99%, suppose now that we are interested in the maximum loss over a shorter horizon, say one month. This is not exactly 1/12 of the previous one, since it must be computed as

$$\begin{aligned} \text{monthly } VaR_{99\%} &= m/12 - 2.326\,3\sigma/\sqrt{12} = \\ &= 15/12 - 2.326\,3 \times 10/\sqrt{12} = -5.46 \text{ millions} \end{aligned}$$

Once VaR has been computed, a variety of potential uses are common. Setting aside VaR on the portfolios of individuals, at the firm level VaR can be used for:

- reporting (measure and control the risk of different financial products and/or firm divisions and/or desks or customers)
- risk-adjusted performance evaluation
- resource allocation (according to a KPMG survey, the 26% of firms in Europe adopt it this way)
- price setting
- construction of risk measures for the whole firm
- setting apart the Capital at Risk (CaR).

In short, it represents a single instrument for a number of different applications.

Let us focus on the last one: CaR is the amount of capital which covers - or is a multiple of - either internally measured VaR or VaR computed according to regulation. These two capitals are named economic and regulatory capital, respectively. CaR is a buffer against potential losses up to a solvency standard, the level of confidence. Suppose now that not enough capital is at disposal in order to cover (or to represent a multiple of) VaR : the firm must reduce its risk exposure, i.e., manage the risk. This is usually done either by insuring some risks or by writing financial contracts. In what follows we analyze the main principles of the two interventions separately.

2 Risk management via insurance contracts

We introduce insurance contract notions, such as refund, premium and loading, in order to give a rationale for the existence of insurance. Namely, we want to demonstrate that it can be profitable for both parties, the insurer (seller) and the insured (buyer, policyholder). In this process, we will refer to a generic loss and claim, not concerning human life. Then, as a typical example of more complex insurance contract, we will analyze life insurance. To conclude, we will introduce the notion of guarantee fund, which backs up the risks insurance companies endorse by selling insurance contracts. This will reconnect us to VaR-like considerations.

2.1 Why insurance exists

Let us consider a basic insurance problem, with a single insurer and a single insured, in which the loss incurred by the policyholder deserves total repayment - here called **refund** - by the insurer (we will subsequently explain why this total coverage is rarely seen in practice). We choose this basic setting, as it appears in Peccati (2002), in order to show that there is room for insurance if its price, the so-called **premium**, is appropriately chosen.

Let us denote by $\mathbf{X}$ the dollar amount claimed by the policyholder; assume that $\mathbf{X}$ can take only two values: S with probability q and 0 otherwise. As a consequence, the distribution of the claim is:

$$\mathbf{X} \sim \begin{cases} \text{values} & \text{probabilities} \\ S & q \\ 0 & 1-q \end{cases}$$

In order to obtain the refund the policyholder pays a price or premium P. It follows that $\mathbf{Y}$, the net cashflow to the insurance company, is:

$$\mathbf{Y} = P - \mathbf{X} \sim \begin{cases} P - S & q \\ & \\ P & 1-q \end{cases}$$

The premium P is called fair if the net expected cash flow is equal to zero (or, equivalently, if it makes inflows equal to the expected value of outflows).

$$\mathrm{E}(\mathbf{Y}) = P - Sq = 0 \Rightarrow P = Sq$$

Actual premia are generally greater than their fair value, and in what follows we will try to understand how much greater they can be, as well as how this relates to the existence of the contract itself. We will then determine their lower bound and upper bound, respectively, as the seller's and buyer's reservation price. It will be assumed that both the seller and buyer are rational decisionmakers à la von Neumann-Morgenstern.

2.1.1 Seller's reservation price

As for the seller, let us denote the insurance company's utility function as u and its initial capital as W_0. The insurer can either sign the insurance contract, in which case its capital becomes stochastic:

$$\mathbf{W}_1 = W_0 + P - \mathbf{X}$$

or not, in which case its capital remains W_0. It will follow the first strategy if and only if

$$u(W_0) \leq Eu(\mathbf{W}_1) = \mathrm{E}u(W_0 + P - \mathbf{X})$$

Since u is always assumed to be increasing in its argument, a lower bound for the premium (insurer's reservation price) is the premium P which makes the above inequality an equality:

$$u(W_0) = \mathrm{E}u(\mathbf{W}_1) = \mathrm{E}u(W_0 + P - \mathbf{X}) \tag{2}$$

Finding a lower bound for the premium amounts to solving equation (2) for P. The result is greater than Sq, provided that the insurance firm is risk averse. It collapses into Sq whenever the insurance firm is risk neutral. In this case its utility function is linear and equation (2) becomes

$$W_0 = \mathrm{E}(W_0 + P - \mathbf{X}) \tag{3}$$

which is solved by $P = \mathrm{E}(\mathbf{X}) = Sq$.

Example (Peccati, 2002) — Suppose for instance that $u(x) = -e^{-kx}$, where k is interpreted as a risk aversion parameter. Equation (2) becomes:

$$-e^{-kW_0} = -(1-q)\,e^{-k(W_0+P)} - qe^{-k(W_0+P-S)}$$

dividing both sides of the equation by $-e^{-kW_0}$, we get:

$$1 = (1-q)\,e^{-kP} + qe^{-kP+kS}$$

and multiplying by e^{kP}:

$$e^{kP} = 1 - q + qe^{kS}$$

taking logarithms and dividing by k:

$$P = \frac{1}{k} \ln \left(1 - q + q e^{qS}\right) \tag{4}$$

Let us also introduce a numerical example to check that P is greater than the fair premium. Set $S = 100$ and $q = 10\%$, so that the fair premium is $100 \times 10\% = 10$.

Suppose that $k = 0.01$. Substituting these figures in (4) we obtain:

$$P = \frac{1}{0.01} \ln \left(0.9 + 0.1 e^{0.01 \times 100}\right) = 15.857$$

which is more than 50% greater than the fair premium.

2.1.2 Buyer's reservation price

Why should a customer agree to pay the insurance company a premium which is not fair? Let us tackle the problem from the policyholder's point of view in order to find his reservation price. Without insurance, he incurs the loss $-\mathbf{X}$. If he signs the insurance contract, he pays P for certain, but he gets rid of the risk Denote with U the insured's utility, with w_0 and $\mathbf{w}_1$ his initial and final wealth respectively. Using the same approach as for the insurer, we can claim that the customer will be willing to buy insurance if and only if

$$\mathrm{E}U\left(w_0 - \mathbf{X}\right) \leq \mathrm{E}\left(U\left(\mathbf{w}_1\right)\right)$$

where

$$\mathbf{w}_1 = w_0 - \mathbf{X} - P + \mathbf{X}$$

The upper premium he is willing to pay is the one which equates the expected utility in the two situations (with and without insurance). It solves

$$\mathrm{E}U\left(w_0 - \mathbf{X}\right) = \mathrm{E}U\left(\mathbf{w}_1\right) = \mathrm{E}U\left(w_0 - \mathbf{X} - P + \mathbf{X}\right) = U\left(w_0 - P\right) \tag{5}$$

The premium which solves (5) is usually greater than the one which solves (2): this happens for instance if the two parties have the same type of utility function, with the buyer more risk averse than the seller, and at the limit the seller risk-neutral. In such a case, the whole interval between the seller′s and bu yer′s reservation prices can be used in order to make at least one of the parties strictly better off than without insurance.

Example — Returning to the numerical example above, suppose that also the customer's utility function is exponential, with a coefficient (of risk aversion) which is twice as large as the insurer's:

$$U\left(x\right) = -e^{-0.02x}$$

The indifference condition (5) becomes:

$$-e^{-0.02(w_0 - P)} = -0.9 e^{-0.02 w_0} - 0.1 e^{-0.02(w_0 - 100)}$$

$$e^{0.02P} = 0.9 + 0.1e^2$$

whose solution is:

$$P = 24.701$$

In this case the policyholder, being more risk averse than the insurer, would accept to pay premia which are much higher than the minimum premium required by the insurer. Every premium between 15.857 and 24.701 makes both parties willing to sign the contract. With a risk neutral insurance company, every premium between 10 and 24.701 would be good.

A risk averse insurer therefore would not apply the fair premium, $\mathrm{E}(X)$, but would start from his lower bound. The departure of the latter from the former (5.857 in the previous example) is called safety margin. This loading refers to departures of actual claims from their expected values, though in reality it also covers differences between q and the actual probability. The cum-safety loading premium is usually tought to be obtained proportionately from the fair one, as follows:

$$P = (1 + \lambda)\,\mathrm{E}(X) \qquad \lambda > 0$$

Furthermore, insurers should cover costs or expenses (administrative, legal, fees, etc...), and can do this by adding a second margin. However, costs can also be explicitly taken into consideration in determining the optimal refund, as illustrated by the following section.

2.1.3 Optimal insurance and deductibles

Full refund is rarely or never seen in practice, due to the fact that, as demonstrated below, it is inconsistent with the existence of variable costs for the insurance company.

Formally, let us consider the fact that the refund can be a function $R(\mathbf{X})$ of the risk, $\mathbf{X}$, instead of coinciding with it. A priori we know that the refund cannot be negative and cannot exceed the risk, $0 \leq R(\mathbf{X}) \leq \mathbf{X}$. The cum-safety margin premium, with refund $R(\mathbf{X})$, will become a functional of refunds:

$$P = (1 + \lambda)\,\mathrm{E}(R(\mathbf{X})) \qquad \lambda > 0$$

Let us also denote as c the costs for the insurance company, and assume that they are a function of the refund : $c(R)$.

The wealth of the insurance company - if the contract is signed - becomes $\mathbf{W}_1 = W_0 + P - R(\mathbf{X}) - c\,[R(\mathbf{X})]$, so that the company will be willing to offer insurance if and only if

$$u\,(W_0) \leq \mathrm{E}u\,(\mathbf{W}_1) = \mathrm{E}u\,(W_0 + P - R(\mathbf{X}) - c\,[R(\mathbf{X})]) \tag{6}$$

Let us imagine that, for a given cost function, the premium and refund policy are chosen so as to maximize the insured's utility, which is now $\mathrm{E}U\,(w_0 - \mathbf{X} - P + R(\mathbf{X}))$, provided that the insurance constraint (6) is satisfied. Formally, let us search for the solutions to the problem

$$\begin{cases} \max_{P,R} \mathrm{E}U\left(w_0 - \mathbf{X} - P + R(\mathbf{X})\right) \\ s.t. \\ 0 \leq R(\mathbf{X}) \leq \mathbf{X} \\ u\left(W_0\right) \leq \mathrm{E}u\left(W_0 + P - R(\mathbf{X}) - c\left[R(\mathbf{X})\right]\right) \end{cases}$$

It turns out (see Raviv (1979)) that the optimal refund is zero up to a given level of damage, proportional to it above:

$$R^*(\mathbf{X}) = \begin{cases} 0 & \mathbf{X} \leq a \\ b\left(\mathbf{X} - a\right) & \mathbf{X} > a \end{cases}$$

with the constants b and a appropriately chosen, as a function of the costs and risk aversion of the insurer. This type of refund, which is the one most seen in practice, is called "with deductible" a. In particular:

- $a = 0$ iff there are fixed costs only ($c' = 0$)
- $b < 1$, which means partial refund over the deductible, either if the insurer is risk averse or if he is risk neutral and costs are convex: $c'' > 0$
- with linear costs and a risk neutral insurer $b = 1$, i.e., there is full refund over the deductible. In this case the payoff of an insurance policy is the same of a European financial call option with strike equal to zero (see section 3 below), and the fair premium is easily computed according to the financial literature on option pricing.

2.2 Life-insurance contracts

In the previous section we referred to generic loss insurance, in order to work with an extremely simple contract. Now that we have shown that there is economic room for insurance, we can proceed with the analysis of some very important types of insurance contracts, related to human life. Our primary focus will be on fair premia.

Let us consider an individual aged x years, also called a life aged x. We denote by $_{/t}q_x$ the probability that a life aged x will die within t years and by $_{/t}p_x = 1 - _{/t}q_x$ the probability that an individual aged x will survive for at least t years. These probabilities are usually computed by taking the number l_{x+t} of survivors after x years, out of an initial number l_x:

$$_{/t}p_x = \frac{l_{x+t}}{l_x}$$

$$_{/t}q_x = 1 - _{/t}p_x = \frac{l_x - l_{x+t}}{l_x}$$

As an example of the latter probabilities, for a life aged $x = 35$ and up to $x + t = 45$, consider the Italian male data in 2000:

t	1	2	3	4	5	6	7	8	9	10
$_tq_{35}$	0.11%	0.22%	0.35%	0.48%	0.61%	0.76%	0.91%	1.08%	1.26%	1.46%

2.2.1 Pure endowment

Let us begin with a contract called pure endowment, which provides for a payment of C dollars to an individual now aged x, if the individual will be alive at the end of the t-th year; otherwise, nothing will be due. We are interested in the present values of the corresponding cash flows. Assuming a constant interest rate (technical rate) equal to i per year, they are $C/\left(1+i\right)^t$ and 0 respectively.

Formally, as in the previous section, the payment of the insurance company is a binary rv:

$$\mathbf{X}\left\{\begin{array}{cc} \text{values} & \text{probabilities} \\ C/\left(1+i\right)^t & {}_{/t}p_x \\ 0 & {}_{/t}q_x \end{array}\right.$$

The so-called expected discounted value of the policy is:

$$\frac{C}{\left(1+i\right)^t} \cdot {}_{/t}\, p_x + 0 \cdot {}_{/t}\, q_x = \frac{C \cdot {}_{/t}\, p_x}{\left(1+i\right)^t}$$

Suppose all l_x individuals aged x in the population sign the same contract with the insurer: they pay a premium and they receive a capital of C after t years if they are still alive. Then, if the premium paid by every insured is equal to the expected discounted value calculated above, the insurer makes zero profits on average. To appreciate this, note that after t years the insurer will pay a capital of C to the l_{x+t} individuals which are still alive, i.e., a total of $C \cdot l_{x+t}$ dollars, whose present value is

$$\frac{C \cdot l_{x+t}}{\left(1+i\right)^t}$$

Face to that, he receives now l_x premia, i.e.,

$$l_x \frac{C \cdot {}_{/t}\, p_x}{\left(1+i\right)^t} = \frac{C \cdot l_{x+t}}{\left(1+i\right)^t}$$

Inflows and outflows are on average (as an effect of risk pooling) the same, and this guarantees that $\frac{C \cdot {}_{/t}p_x}{(1+i)^t}$ is the fair risk premium.

As an example, consider a pure endowment of $C = 100$ in 10 years ($t = 10$), for a life aged 35, with an interest rate $i = 2\%$. Based on the above table, $_{10}p_{35} = 1 - 1.46\% = 98.54\%$, so that its fair premium is

$$P = 100 \times 98.54\% \times 1.02^{-10} = 80.84$$

2.2.2 Annuity

Having explained how to value an endowment, it become easy to value a life annuity, which entitles a policy holder aged x to receive C in one year if he is alive then, C in two years if he is alive then, and so forth. The value of the life annuity is obtained by summing the expected discounted values of the single payments:

$$\frac{C \cdot_{/1} p_x}{(1+i)} + \frac{C \cdot_{/2} p_x}{(1+i)^2} + \frac{C \cdot_{/3} p_x}{(1+i)^3} + \cdots + \frac{C \cdot_{/110-x} p_x}{(1+i)^{110-x}}$$

We stop summing terms at the age of 110, under the assumption that nobody can survive longer: $l_{110} = 0$. The above sum represents the fair premium P for an annuity. It also represents the fair value of a (guaranteed) pension, independently of the pillar.

2.2.3 Term insurance and whole life

Let us now consider term insurance contracts, which pay off a capital C at the end of the year of death of the insured, if it occurs within k years from now. As a limit of term insurance, when k extends to $110 - x$, we will also consider the so-called whole-life policies.

Note first of all that the probability that a life aged x will survive for t years and then die within the end of the $(t+1)$-th year is:

$$_{/t}p_x \ \cdot_{/1} q_{x+t}$$

i.e., it is equal to the product of the probability of surviving for t years and the (conditional) probability of dying within one year after having reached the age $x + t$. The expected discounted value of the payment due if the insured dies during the $(t+1)$-th year is:

$$_{/t}p_x \ \cdot_{/1} q_{x+t} \cdot \frac{C}{(1+i)^{t+1}}$$

Since the previous guarantee is extended to k years, we have as an expected discounted payment (and therefore fair premium P):

$$P =_{/1} q_x \cdot \frac{C}{(1+i)^1} +_{/1} p_x \ \cdot_{/1} q_{x+1} \cdot \frac{C}{(1+i)^2} +_{/2} p_x \ \cdot_{/1} q_{x+2} \cdot \frac{C}{(1+i)^3} +$$

$$+ \ldots +_{/k-x} p_x \ \cdot_{/1} q_{x+k} \cdot \frac{C}{(1+i)^{k+1}}$$

Let us consider a term insurance lasting $k = 3$ years, subscribed by an Italian male in 2000, with a technical rate equal to $i = 2\%$. Since

$_{/1}q_{35}$	$_{/1}p_{35} \cdot_{/1} q_{36}$	$_{/2}p_{35} \cdot_{/1} q_{37}$
0.11%	(1-0.11%)·0.22%	(1-0.35%)·0.12%

the fair premium is

$$P = 0.11\% \cdot \frac{100}{1.02} + (1 - 0.11\%) \ \cdot 0.22\% \cdot \frac{100}{1.02^2} + (1 - 0.35\%) \cdot 0.12\% \cdot \frac{100}{1.02^3}$$

If the insurance is a whole-life one, namely if it extends to every year from 1 to $110 - x$, the value of the policy is given by the following sum:

$$_{/1}q_x \cdot \frac{C}{(1+i)^1} +_{/1} p_x \ \cdot_{/1} q_{x+1} \cdot \frac{C}{(1+i)^2} +_{/2} p_x \ \cdot_{/1} q_{x+2} \cdot \frac{C}{(1+i)^3} +$$

$$+ \ldots +_{/110-x} p_x \ \cdot_{/1} q_{x+110} \cdot \frac{C}{(1+i)^{110+1}}$$

which represents also the policy's fair premium P.

2.2.4 Mathematical Reserves

Insurance companies must report in their financial statements the difference between the value of their obligations towards the policyholders and the value of the (fair) premia to be received from them, the so called mathematical reserve[4].

Since fair premia are chosen so as to equate expected future payments and future benefits, the reserve initial level, before any premium is paid, is zero.

For instance, if a company has issued a single pure endowment policy, with a single premium, its reserve, just before the premium payment, is

$$-P + \frac{C \cdot_{/t} p_x}{(1+i)^t} = 0$$

while the reserve just after the premium payment becomes

$$\frac{C \cdot_{/t} p_x}{(1+i)^t}$$

After one year, the reserve becomes

$$\frac{C \cdot_{/t-1} p_{x+1}}{(1+i)^{t-1}}$$

[4]The hypothesis made thus far is that the premium is entirely paid by the insured when the contract is signed. However, it is quite common that the premium is paid in installments; installments in turn are paid by the insured only if he is alive. Let us denote as P_t the premium installment to be paid at time t (P_1 in one year, P_2 in two years, P_3 in three years and so on), and with n the total number of installments. The premium P_t will be paid with probability $_tp_x$.

To preserve actuarial fairness the amount of the installments must be such that their expected discounted value is equal to the single premium that would be paid when the contract is signed:

$$P =_{/1} p_x \cdot \frac{P_1}{1+i} +_{/2} p_x \cdot \frac{P_2}{(1+i)^2} +_{/3} p_x \cdot \frac{P_3}{(1+i)^3} + \ldots +_{/n} p_x \cdot \frac{P_n}{(1+i)^n}$$

As an example, we can consider the pure endowment above. We have a capital insured $C = 100$ in $t = 10$ years, an interest rate $i = 2\%$: a head aged $x = 35$ years deserves a mathematical reserve at the inception of the contract, but after the premium, equal to 80.84, which becomes

$$\frac{C \cdot {}_{/9}\, p_{36}}{(1+i)^9} = \frac{100 \cdot 98.65\%}{1.02^9} = 82.55$$

after one year.

2.3 Guarantee funds

National laws in various countries are targeted towards the reduction of the risk that the insurance company may not be able to meet its obligations towards the insured. In order to attain this objective, they force insurance companies to set aside so-called 'guarantee funds'. How is the minimum guarantee fund determined? We will follow the brief and -in our opinion - clear exposition in Peccati (2002).

Consider an insurance company that has sold N insurance policies: she has received the premia $P_1, P_2, \ldots, P_N$, for a total of $P = P_1 + P_2 + \ldots + P_N$ dollars, and is liable to pay the random amounts of money $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$, whose sum is $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_N$. Denote the amount of the actual guarantee fund by G, as opposed to the minimum guarantee, G'.

The minimum $G'$is fixed with the help of the so called risk theory, as follows.[5]

Denote by $m_1, m_2, \ldots, m_N$ the expected values of the single payments due by the insurance company $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$ and by $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ their variances. The expected value m of the total (random) amount of payments, $\mathbf{X}$, is equal to the sum of the expected values of the single payments:

$$m = m_1 + m_2 + \ldots + m_N$$

We will assume that the payments are stochastically independent: the variance σ^2 of $\mathbf{X}$ therefore equals the sum of the variances of the payments

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_N^2$$

A reasonable way of selecting the minimum level G' of the guarantee fund consists in choosing it in such a way that the ruin probability, i.e.,, the probability that the insurance company goes bankrupt, be small. Let us denote by α the ruin probability: α is equal to the probability that the sum P of the premia paid by the insured and the guarantee fund G' are not sufficient to meet the total amount of payments $\mathbf{X}$

$$\Pr\left(G' + P - \mathbf{X} < 0\right) = \alpha \qquad (7)$$

[5] [4] compares risk theory with previously adopted rules.

The inequality in parentheses can be rewritten as:

$$\mathbf{X} > G' + P$$

or

$$\frac{\mathbf{X} - m}{\sigma} > \frac{G' + P - m}{\sigma}$$

At the core of risk theory lies one of the fundamental theorems of probability: the central limit theorem, whose thesis is about the distribution of the standardized rv $\mathbf{Y} = (\mathbf{X} - m)/\sigma$. With a slight lack of rigor, the theorem states that :

- if the number n of policies is large enough (this can be assumed to be true for large insurance companies);
- if the payments are stochastically independent (as assumed);
- if the order of magnitude of some payments is not much larger than that of the other payments (in which case the sum of the payments would be well approximated by the sum of those few payments whose order of magnitude is larger);

then the probability distribution of $\mathbf{Y}$ is approximately standard normal, i.e., $\mathbf{Y}$ has the distribution function in (1).

As a consequence the ruin probability is:

$$\begin{aligned}
\Pr\left(G' + P - \mathbf{X} < 0\right) &= \Pr\left(\mathbf{Y} > \frac{G' + P - m}{\sigma}\right) = \\
&= 1 - \Pr\left(\mathbf{Y} \leq \frac{G' + P - m}{\sigma}\right) = \\
&= 1 - \Phi\left(\frac{G' + P - m}{\sigma}\right)
\end{aligned}$$

The guarantee fund G' then has to be chosen in such a way that:

$$1 - \Phi\left(\frac{G' + P - m}{\sigma}\right) = \alpha$$

Solving with respect to (from now on, wrt) G' we obtain:

$$\frac{G' + P - m}{\sigma} = \Phi^{-1}(1 - \alpha)$$

$$G' = m + \sigma\Phi^{-1}(1 - \alpha) - P \tag{8}$$

which can be read as: "The minimum margin must cover the difference between the VaR of the insurance company and its premia".

As in the VaR case, the value $\Phi^{-1}(1 - \alpha)$ can be read from the statistical tables: if $\alpha = 1\%$ for instance it is 2.33.

The formula for G' is then very easy to implement: combining the value of $\Phi^{-1}(1-\alpha)$ with the characteristics of the portfolio of insurance policies, i.e., m, P and σ, it is straightforward to calculate the minimum amount of the guarantee funds to be held. Suppose for instance that an insurance company has issued $N = 10,000$ policies, with mean refund $m = 50,000\$$, standard deviation $\sigma = 2,500\$$, and has received a total premium $P = 15,000$. Its minimum guarantee fund, in order to have a ruin probability of 1%, is $G' = 50,000 + 2,500 \times 2.33 - 15,000 = 40,825$

Let us now consider the actual amount G of funds, instead of the minimum amount G': the ratio

$$\frac{G + P - m}{\sigma}$$

is called solvency ratio. If we refer to the formulas listed above, it becomes apparent that the higher the solvency ratio, the less likely it is the the insurance company will go bankrupt. As a consequence, the ruin probability decreases when the premia P or the guarantee fund G increase, and when the expected value of the payments m or their standard deviation (variability) σ decrease. If the solvency ratio is too low, the possible remedies are an increase in the premium paid by the subscribers of new policies or the reduction of refunds to them. It is also possible to reduce σ, by reinsuring some of the risks.

3 Risk management via financial contracts

Another opportunity for reducing risks, such as market and credit ones, consists in trading in financial securities, especially derivative ones.

Derivatives are securities whose price depends on the price of another asset, called underlying asset, usually a stock or bond or exchange rate. Let us denote the underlying price with $\mathbf{S}$, the derivative one as $\mathbf{Y}$. Since the latter is a function f of $\mathbf{S}$, we will write $\mathbf{Y} = f(\mathbf{S})$.

Consider, for example, a European call option on a stock: the option gives its holder the right (but not the obligation) to buy one share of the stock for a fixed price E, called strike price, at a fixed maturity (called expiration). If at the expiry date the price $\mathbf{S}$ of the stock is greater than the strike price E, the option is worth exercising, otherwise it is not. The value of the option is therefore positive and equal to the difference between $\mathbf{S}$ and E in the first case, it is zero in the second:

$$\mathbf{Y} = f(\mathbf{S}) = \begin{cases} \mathbf{S} - E \text{ if } \mathbf{S} > E \\ \\ 0 \text{ otherwise} \end{cases}$$

In this chapter the issue of trading derivatives - or hedging via derivatives in order to reduce risk - will be strictly linked to the problem of fairly pricing such derivatives. We start by providing some basic notions of derivative pricing in what is referred to as binomial framework (i.e. when the underlying at maturity can take only two values). In this process, it will be assumed that financial markets are free from arbitrage opportunities or, in other words, of opportunities to earn money for certain and at no cost. This is the standard assumption in the field.

3.1 Derivative pricing

Let us stick to the European call example and consider the case in which, at the point of maturity (i.e. one month) the underlying can have two possible values only, s_1 or s_2, respectively smaller and bigger than today's value of the stock:

$$s_1 < S_0 < s_2$$

At expiry, the value of the option will be either $y_1 = f(s_1)$ or $y_2 = f(s_2)$.

If we want to determine the no-arbitrage price of the option today, keeping in mind insurance pricing, the first idea which comes to mind is:

1. assess the probabilities of the two possible prices: $p_1 = \Pr(\mathbf{S} = s_1)$ and $p_2 = \Pr(\mathbf{S} = s_2)$,
2. note that these probabilities are also the ones of the two possible values of the derivative security: $p_1 = \Pr(\mathbf{Y} = y_1)$ and $p_2 = \Pr(\mathbf{Y} = y_2)$,
3. calculate the expected value of $\mathbf{Y}$:

   $$\mathrm{E}(\mathbf{Y}) = y_1 p_1 + y_2 p_2$$

   and, remembering that the payoff of the derivative will take place in one month,
4. discount it at the monthly rate r:

   $$w = \frac{y_1 p_1 + y_2 p_2}{1 + r} \tag{9}$$

This valuation procedure does not give the correct no-arbitrage price of the derivative, since it allows for arbitrage possibilities Consider for instance the case in which $s_1 = 90, s_2 = 110, E = 100$. We have:

$$\mathbf{Y} = f(\mathbf{S}) = \begin{cases} y_2 = 10 \text{ if } \mathbf{S} = s_2 = \mathbf{110} > E = 100 \\ \\ y_1 = 0 \text{ if } \mathbf{S} = s_1 = \mathbf{90} < E = 100 \end{cases}$$

Suppose also that $r = 2\%, S_0 = 95$, and the probabilities are both 50%. We get

$$w = \frac{y_1 p_1 + y_2 p_2}{1 + r} = \frac{10 \times 50\%}{1.02} = 4.9 \tag{10}$$

At this price, an arbitrage opportunity consists in:

- selling two options for 4.9,
- buying the underlying for 95 and
- borrowing the present value of 90, i.e. 90/1.02.

This gives a total, positive inflow of +4.9×2-95+98.04>0 at time 0.

After one month

- if $\mathbf{S} = s_2 = \mathbf{110}$, the options are exercised by the buyer, so that the arbitrageur has an outflow of -20. At the same time, he owns the underlying, which is worth 110, and has to repay the compound amount of 90/1.02, i.e., 90. On the whole, his cashflow is -20+110-90=0.

- if $\mathbf{S} = s_1 = \mathbf{90}$, the options are notexercised by the buyer, so that the arbitrageur has no outflow. The underlying is worth 90, and has to repay the compound amount of 90/1.02, i.e., 90. On the whole, his cashflow is 0+90-90=0.

The reader can check that he obtains the same cashflow in case $\mathbf{S} = s_1$: there has been an arbitrage indeed, with a profit at time 0 and no flow (for sure) at expiry.

In order to find the correct option price, we are left with no other technique than ruling out arbitrage possibilities.

In particular, let us consider the arbitrage possibilities stemming from a portfolio of the underlying and the derivative. Let the portfolio be formed by one unit of the derivative security and h shares of the stock: in technical jargon, the portfolio is long the derivative and long or short the underlying, according to whether h is positive or negative. At maturity the portfolio value will be

$$\begin{cases} y_2 + hs_2 \text{ if } \mathbf{S} = s_2 \\ \\ y_1 + hs_1 \text{ if } \mathbf{S} = s_1 \end{cases}$$

If we choose h in such a way that

$$y_1 + hs_1 = y_2 + hs_2$$

i.e., if

$$h^* = \frac{y_1 - y_2}{s_2 - s_1} = -\frac{f(s_2) - f(s_1)}{s_2 - s_1}$$

then, independently of the final stock price, the final value of our portfolio will be the same: the portfolio itself will be riskless, and therefore deserve the return r.

Let us use this remark in order to find the correct derivative price, v. Today, the portfolio is worth:

$$v + h^* S_0 = v + \frac{y_1 - y_2}{s_2 - s_1} S_0$$

and it will give $y_1 + h^* s_1$(or $y_2 + h^* s_2$) for sure tomorrow. The latter amount must represent the compound amount of the initial investment, at the rate r:

$$\left(v + \frac{y_1 - y_2}{s_2 - s_1} S_0 \right)(1 + r) = y_1 + \frac{y_1 - y_2}{s_2 - s_1} s_1$$

This no arbitrage-relationship is a first degree equation in the unknown v, whose solution yields the no-arbitrage price of the derivative:

$$v = \frac{1}{1+r}\left[y_1 \frac{s_2 - S_0(1+r)}{s_2 - s_1} + y_2 \frac{S_0(1+r) - s_1}{s_2 - s_1}\right] \tag{11}$$

In the above numerical example, since $S_0 = 95$, we have

$$v = \frac{1}{1.02}\left(10\frac{95 \times 1.02 - 90}{110 - 90}\right) = 3.38$$

3.2 Risk-neutral probabilities

Please notice that the no-arbitrage price (11) can be re-written as:

$$v = \frac{1}{1+r}(y_1 q_1 + y_2 q_2) \tag{12}$$

if one posits

$$q_1 := \frac{s_2 - S_0(1+r)}{s_2 - s_1},\ q_2 := \frac{S_0(1+r) - s_1}{s_2 - s_1}$$

The "quantities" q_1 and q_2 have the following properties:

- they are positive, provided that[6]

$$s_1 < S_0(1+r) < s_2$$

- their sum is equal to 1

In the example, we have

$$q_1 = \frac{110 - 95 \times 1.02}{110 - 90} = 0.655, q_2 = \frac{95 \times 1.02 - 90}{110 - 90} = 0.345$$

We can thus interpret v as the discounted value of the expectation of the future payoff of the derivative, where the expected value has not been calculated using the true probabilities p_1, p_2, but the mock probabilities q_1, q_2 (also called synthetic). The latter depend only on the future price of the stock, its current price and the risk-free rate r. The correct price of the derivative security is equal to a discounted expectation, as firstly envisaged, but with a strong proviso on the probabilities.

[6]This in turn is necessary for the derivative market to exist: with $s_2 > s_1 \geq S_0(1+r)$ in fact nobody would buy the riskless asset, which would be dominated by the stock; with $S_0(1+r) \geq s_2 > s_1$ nobody would buy the stock, which would be dominated.

3.3 Hedging

As a byproduct of the pricing technique illustrated above, we remarked that the portfolio composed of one derivative and

$$h^* = -\frac{f(s_2) - f(s_1)}{s_2 - s_1}$$

underlyings is risk-free, in the sense that it does not change value whenever the underlying does. Let us denote $f(s_2) - f(s_1)$ as Δy, $s_2 - s_1$ as ΔS : the hedge ratio, i.e., the number of underlying contracts which has to be held, together with one (long) unit of derivative, in order to be protected from risk, or insured against price risk, in the binomial model, is $h^* = -\Delta y/\Delta S$. In the example, $h^* = -10/20 = -.5$, i.e., selling half an underlying (or an underlying each two derivative contracts) is necessary. With this hedge, if state 1 occurs, the portfolio is worth 0-90/2=-45, while if state 2 occurs it is worth 10-110/2=-45.

Generally speaking, whenever

$$y = f(S) \tag{13}$$

any change of the price S of the underlying asset, ΔS, induces a change in the price of the derivative, Δy. The ratio $\Delta y/\Delta S$ tells us by how many dollars the price of the derivative changes, when the price of the underlying asset changes by one dollar. The reason why a portfolio of the type

$$P = y + h^* S = y - \frac{\Delta y}{\Delta S} S$$

is risk-free is that, by the choice of h^*

$$\Delta P = \Delta y - \frac{\Delta y}{\Delta S} \Delta S = 0$$

More generally, if prices can move continuously, we take as optimal ratio h^* the limit of the ratio $\Delta y/\Delta S$ for ΔS which tends to zero, i.e.,:

$$\lim_{\Delta S \to 0} \frac{\Delta y}{\Delta S} = \lim_{\Delta S \to 0} \frac{f(S + \Delta S) - f(S)}{\Delta S} = \frac{\mathrm{d}f}{\mathrm{d}S} = f'(S)$$

This limit is called Delta of the derivative asset and is denoted by δ. As in the discrete case, starting from a portfolio

$$P = y - \delta \times S$$

and recalling the notion of a differential, we can write down the variation in the value of the portfolio induced by a small variation in the price of the underlying asset:

$$\begin{aligned} \mathrm{d}P &= \mathrm{d}(y - \delta \times S) = \\ &= \mathrm{d}(f(S) - \delta \times S) = \\ &= f'(S)\,\mathrm{d}S - \delta \times \mathrm{d}S \end{aligned} \tag{14}$$

The variation is zero since $\delta = f'(S)$.

This technique, widely adopted in portfolio management, is called Delta-hedging: one eliminates the riskiness of a portfolio by deciding its composition in such a way that the changes in value of the single assets exactly offset each other. Delta are also required in order to compute the VaR of portfolios with derivatives and are asked for by regulation Authorities in their surveillance reports.

4 Agency problems in contract fulfilling: action risk

In this section we will analyze the concept of moral hazard, following Kreps (1990) and Tirole (1998). This topic envisages a typical problem of contract fulfilling, which arises any time the action of the "agent" in the contract cannot be enforced by the "principal". Loosely speaking, moral hazard is the risk that the agent will not act "optimally" for the principal. While financial risks are hedged via insurance or derivatives, moral hazard is hedged via incentives for the agent, to be specified in the design of the contract.

Moral hazard will therefore lead us to the issue of optimal contract design and its mathematical modelling.

Another common problem, which can be dealt with via contract design and can be formally treated with similar mathematical tools, is the so-called adverse selection. The impossibility of assessing the quality of a product or of a counterpart squeezes out high quality goods and counterparts. As with moral hazard, sale contracts need to be enriched with provisos (guarantees, in this case) in order to mitigate the adverse selection phenomenon. We simply touch upon the similarity of the problem in the present context, and refer the reader to Baron (1986), Kreps (1990) for further reading.

As a basic example of moral hazard, we will consider the interaction between a risk-neutral firm, which acts as a principal, p, and a unique risk-averse manager, the agent a. The firm and its shareholders are seen as a single entity, so that we will refer to them interchangeably. Both the principal and the agent aim at maximizing the expected utility of net income, which is represented by the firm's profits minus the manager's wages for the firm, and by wages minus the monetary value of effort for the the manager. The expectation is taken with respect to business or external conditions which are, generally speaking, represented by a random variable X. In the firm scenario, given risk neutrality, maximizing net income and its utility is the same process.

The firm receives profits π, which are assumed to depend both on X and on the manager's effort e: $\pi = \pi(e, X)$. It pays the manager's wage, which in turn is assumed to depend on the profits earned: $w = w(\pi)$. Therefore, the firm's net income or profit is

$$\pi(e, X) - w(\pi(e, X))$$

For each effort level e^*, the firm maximizes expected net profit, by appropriately choosing a payment scheme for its agent, namely a function w

$$\max_{w} E_X \left[\pi(e^*, X) - w(\pi(e^*, X)) \right]$$

As far as the manager is concerned, his utility is defined on the difference between the wage he receives and his effort, transformed into monetary units through a linear factor $k : U = U(w - ke)$, $k > 0$. Let us assume $U' > 0, U'' < 0$.

First of all, the firm must check that under the chosen wage function, the agent employs exactly the level of effort she subsumed when computing that wage (incentive compatibility constraint):

$$e^* = \arg\max_{e} E_X U \left[w(\pi(e, X)) - ke \right]$$

In addition, the manager will work if and only if his utility by participating in the labour market is greater than the one attained by not working. If we denote the latter as U_0, the firm will have to reach its own aim under the following constraint (participation constraint):

$$U(w - ke^*) \geq U_0$$

To sum up, the firm-principal will have to solve the following problem:

$$\mathcal{P}_p = \begin{cases} \max_w E_X \left[\pi(e^*, X) - w(\pi(e^*, X)) \right] \\ s.t. \\ U(w - ke^*) \geq U_0 \\ e^* = \arg\max_e E_X U \left[w(\pi(e, X)) - ke \right] \end{cases}$$

We will analyze this problem at increasing levels of complexity: first of all (subsection 1), we will imagine that there is no business or external uncertainty, that the principal is totally inactive, and that the agent decides his level of effort. We will notice that in such a framework the aims of the two parties can be reached at the same time if and only if the manager is also the firm's owner.

Given the conflict between the two parties when they do not coincide, we will define a decisional problem for the principal which aims at circumventing such conflicts. We will solve the problem armed with full knowledge of the manager's effort (section 2), as well as without the possibility of observing such effort (section 3).

4.1 Conflicting aims: the moral hazard origin

In order to make the reader aware of the conflict between the two parties, and how it surfaces formally, let us imagine that

- no external uncertainty exists ($X = 0$ wp1), so that profits depends on the agent effort only: $\pi(e, X) = \pi(e)$
- the agent is risk-neutral and has reservation utility equal to zero ($U_0 = 0$)

- all the action is left to the agent
- the firm does not decide a wage function, but wage is simply proportional to the profit: $w(\pi(e,X)) = b\pi(e)$, with $b \in \Re^+$.

It is easy to demonstrate that, even under these simplifying assumptions, the two parties cannot maximize their utility at the same time. Let us write down the problem for the agent:

$$\begin{aligned} \max_e [w(\pi(e)) - ke] &= \max_e [b\pi(e)) - ke] \\ & \quad s.t. \\ b\pi(e) - ke &\geq 0 \end{aligned}$$

The first order condition (from now on, foc) for an internal maximum is:

$$b\pi'(e) - k = 0 \tag{15}$$

Even if the effort which satisfies this condition, $e^* = \pi'^{-1}(k/b)$, also satisfies the participation constraint, $b\pi(e^*) - ke^* \geq 0$,the firm's net revenue (and therefore its utility), namely

$$\pi(e) - w(\pi(e)) = \pi(e)(1-b)$$

is not maximized. The foc for its maximum wrt e would in fact be

$$(1-b)\pi'(e) = 0 \tag{16}$$

which is consistent with the manager's one, (15), if and only if $b = 1$, i.e., if the entire profit goes to the manager. In the present case the manager's coincides with the shareholders'. We have formalized the fact that, if the two parties are really distinct ($b \neq 1$), conflicts of interest make the fulfilment of a contract by one of the parties suboptimal for the other. In the next section we will return to the setting in which the firm chooses a wage function which can provide an incentive for the manager to act optimally in his own interest (incentive & compatibility constraints), while at the same time maximizing the shareholders' utility. We will do this initially by assuming that the manager's effort is observable, and subsequently by considering a case in which it is not.

4.2 Conflicting aims & full information

The way out of the impasse described above consists in designing the fee for the agent so as to invite him to participate, while optimizing both his own and the shareholders' net revenue. Suppose that the effort can be observed, so that information is complete. Let us also suppose, for the sake of simplicity, that the rv X, which represents business conditions, can only take the values x_1 and x_2, with probabilities p_1 and p_2.

The firm has to solve the problem

$$\mathcal{P}_p = \begin{cases} \max_w \mathrm{E}\left[\pi - w\right] = \\ \max_w p_1 \left[\pi(e^*, x_1) - w(\pi(e^*, x_1))\right] + p_2 \left[\pi(e^*, x_2) - w(\pi(e^*, x_2))\right] \\ s.t. \\ e^* = \arg\max_e p_1 U \left[w(\pi(e, x_1)) - ke\right] + p_2 U \left[w(\pi(e, x_2)) - ke\right] \\ p_1 U \left[w(\pi(e^*, x_1)) - ke^*\right] + p_2 U \left[w(\pi(e^*, x_2)) - ke^*\right] \geq U_0 \end{cases}$$

Suppose for the time being that the effort level cannot be chosen: e is given. Call w_i the wage that will be given in the business condition i, with effort $e :$

$$w_i := w(\pi(e, x_i))$$

Since there are only two states of the world, choosing a function w_i means choosing two (positive) constants w_1, w_2. The firm problem becomes

$$\mathcal{P}_p \mid e = \begin{cases} \max_{w_1 w_2} p_1 \left[\pi(e, x_1) - w_1\right] + p_2 \left[\pi(e, x_2) - w_2\right] \\ s.t. \\ p_1 U \left(w_1 - ke\right) + p_2 U \left(w_2 - ke\right) \geq U_0 \end{cases}$$

where $\mathcal{P}_p \mid e$ is the principal problem, given the action e.In this case the Lagrangian is

$$L(w_1, w_2, \lambda) = p_1 \left[\pi(e, x_1) - w_1\right] + p_2 \left[\pi(e, x_2) - w_2\right] +$$

$$+\lambda \left[p_1 U \left(w_1 - ke\right) + p_2 U \left(w_2 - ke\right) - U_0\right]$$

and its foc[7] are

$$\begin{cases} -1 + \lambda U'(w_i - ke) = 0 \qquad\qquad i = 1, 2 \\ \lambda \geq 0, w_i > 0 \\ p_1 U \left(w_1 - ke\right) + p_2 U \left(w_2 - ke\right) - U_0 \geq 0 \\ \lambda \left[p_1 U \left(w_1 - ke\right) + p_2 U \left(w_2 - ke\right) - U_0\right] = 0 \end{cases}$$

Analyzing the conditions

$$-1 + \lambda U'(w_i - ke) = 0 \qquad\qquad i = 1, 2$$

one realizes that there cannot be a solution with $\lambda = 0$, and consequently the participation constraint must be satisfied as an equality:

$$p_1 U \left(w_1 - ke\right) + p_2 U \left(w_2 - ke\right) = U_0 \tag{17}$$

In addition, the marginal utility of wage must be the same across states:

$$U'(w_i - ke) = \frac{1}{\lambda} \tag{18}$$

This happens if and only if wages are constant across business conditions, i.e., $w_1 = w_2$: all the firm risk must be borne by the shareholders, and the manager should not participate in it. In this process, as evidenced by (17) above,

[7]Please note that whenever the program is concave, the FOC are sufficient.

shareholders reduce the manager to its reservation wage w^*, namely the one for which the participation constraint is binding:

$$p_1 U\left(w^* - ke\right) + p_2 U\left(w^* - ke\right) = U_0$$

To sum up, $w_1 = w_2 = w^*$: shareholders must give the manager full insurance and, conversely, enjoy all the profit.

What happens if the level of effort is observable and can be chosen? Suppose that the effort can be either positive, h, or zero, 0. Firms choose the wage as above for each level of effort, by solving both $\mathcal{P}_p \mid h$ and $\mathcal{P}_p \mid 0$: call the two reservation wages $w^*(h)$ and $w^*(0)$ respectively. Then, they compare their own utility under $w^*(h)$ and $w^*(0)$ to decide whether they are better off with an high or low effort:

$$p_1\left[\pi(h, x_1) - w^*(h)\right] + p_2\left[\pi(h, x_2) - w^*(h)\right] \geq$$

$$p_1\left[\pi(0, x_1) - w^*(0)\right] + p_2\left[\pi(0, x_2) - w^*(0)\right]$$

If their preferred effort is h, they allow the manager to have $w^* = w^*(h)$ if he exerts the effort h, otherwise 0. If their preferred effort is 0, they give him $w^*(0)$ if he exerts zero effort, otherwise he receives nothing. Since they have the power to threaten him, the problem they have to solve must not be revised by inserting the incentive compatibility constraint in it. When full information is available on the effort level, the firm simply solves $\mathcal{P}_p \mid h$ and $\mathcal{P}_p \mid 0$, as defined above, and consequently forces the manager to act according to its will by choosing the wage and the penalty (0 wage).

In order to extend the discussion to the case in which the effort is not observable, suppose that in our setting it is optimal to induce a high level of effort.

4.3 Conflicting aims and asymmetric information

We are now ready to analyze the complete and more realistic setup of moral hazard, in which effort is unobservable, information is asymmetric, and therefore threats are of no value. We cannot envisage a function $w(h)$ or $w(0)$, since shareholders do not observe the effort, but only a specific profit value. At most, they can make wage a function of the observable result, namely profits: $w = w(\pi)$. We also suppose that the likelihood of a single profit is different under different efforts. Formally, there are two different probability distributions for profits, according to high and low effort, which we denote as $F(\pi \mid h), F(\pi \mid 0)$ respectively. For the sake of this example, let us assume that profits can take n levels, $\pi_i, i = 1, ..n$ (n was equal to 2 in the previous section), put in increasing order, so that $\pi_1 < \pi_2 < ... < \pi_n$.Denote their probabilities under the two distributions as p_i^h, p_i^0 respectively.

Returning to our assumption in the previous section, shareholders know that high effort is preferable, and should devise a way to induce the agent to exert it and to participate. They want his expected utility with high effort to be larger than in low effort cases (incentive-compatibility constraint):

$$\sum_{i=1}^{n} p_i^h U\left(w(\pi_i) - kh\right) \geq \sum_{i=1}^{n} p_i^0 U\left(w(\pi_i)\right) \tag{19}$$

and they want him to participate, given that he exerts the high level of effort:

$$\sum_{i=1}^{n} p_i^h U\left(w(\pi_i) - kh\right) \geq U_0 \tag{20}$$

Under an high effort, their own expected utility is

$$\mathrm{E}\left[\pi - w\right] = \sum_{i=1}^{n} p_i^h \left(\pi_i - w(\pi_i)\right)$$

and maximizing it wrt the w function is equivalent to minimizing

$$\sum_{i=1}^{n} p_i^h w(\pi_i)$$

As in the 2-states case, choosing a function w here means choosing its values in the n states of the world: $w(\pi_i), i = 1, ..n$.

Let the U function be separable in wage and effort, $U = u(w) - ke$: let us also denote as y_i the wage-based utility reached by the agent when profits take their i value, and therefore wage is $w(\pi_i)$. We have, by definition, $y_i := u(w(\pi_i))$. To complete the model, assume that u is strictly increasing and denote as v the inverse of u: by the definitions of v and y_i, the wage which is paid to the agent when the profit is $\pi_i, w(\pi_i)$, is such that

$$w(\pi_i) = v\left(y_i\right)$$

and the corresponding utility for the agent is

$$U = u(w(\pi_i)) - ke = u(v(y_i)) - ke = y_i - ke$$

Let us re-write the optimization problem of the firm wrt the variables y_i :

$$\mathcal{P}_p = \begin{cases} \min_{y_i} \sum_{i=1}^{n} p_i^h v(y_i) \\ s.t. \\ \sum_{i=1}^{n} p_i^h y_i - kh \geq U_0 \\ \sum_{i=1}^{n} p_i^h y_i - kh \geq \sum_{i=1}^{n} p_i^0 y_i \end{cases}$$

This problem can be solved by a Lagrangian formulation. Call its value function, the minimized value of wages, $C(h)$:

$$C(h) := \sum_{i=1}^{n} p_i^h v(y_i^*)$$

where y_i^* are the optimal values of the y_i variables.

In general, both the value function under the high level of effort and the one under the low one, $C(0)$, must be computed and their absolute level compared in order to solve the original problem. The typical features of the solution is the following inequality, which holds for all i:

$$\frac{p_i^h}{p_i^0} \geq \frac{p_{i-1}^h}{p_{i-1}^0}$$

Recalling that the $i-$ profit is by assumption greater than the $i-1$- one, the inequality says that the relative likelihood of the better profit, p_i^h/p_i^0, is greater than the same likelihood for a worse profit, p_{i-1}^h/p_{i-1}^0. Since shareholders at this point cannot directly observe their manager's effort, they cannot force him to exert a specific effort. The best they can do is the increase the likelihood of high profits. This structure nicely applies to more than two (actually, to a continuum of) effort levels and profits.

As we will illustrate in Section (5.2) below, when the agent is risk-neutral all the action risk must be borne by him so that, as opposed to the full information case, the principal has full insurance.

5 Agency problems in litigation

5.1 INSURANCE FRAUD AND OPTIMAL CLAIM SETTLEMENT STRATEGIES (K.J. Crocker and S. Tennyson)

Liability insurers face a principal-agent problem, in that their claimants (agents) can permanently misrepresent losses and, by engaging in claims falsification, file a claim higher than the actual damage suffered. Claims falsification is costly: the agent's effort now becomes a monetary cost. The incentive, on the other hand, must be included in the optimal indemnification strategy, fixed by the principal-insurer.

Crocker and Tennyson (2002) demonstrate that the optimal indemnification involves systematic underpayment of claims at the margin, limited by expected litigation costs and potential bad-faith claims, unless it is extremely costly for policyholders to claim more than the actual loss.

Their theoretical model considers a non-risk-neutral claimant and a risk-neutral insurer. Let us denote the utility of the former with U, as above, and assume $U' > 0, U'' < 0$.

He has suffered an injury whose magnitude x is private information (which corresponds to the effort in the firm-manager example). He files the claim y ($\geq x$), which, as opposed to the injury magnitude, is made known to the insurer-principal. Filing a claim is done under a cost $f(x-y)$, which is assumed to be quadratic: $f(x-y) = \xi(x-y)^2$, with ξ a positive constant. It is therefore increasingly costly to engage in falsification. We also assume that no cost is incurred by "telling the truth", i.e., filing exactly $x : f(0) = 0$.

The indemnity w received face to the claim y is assumed to be a linear function of the request y :

$$w(y) = a + by$$

The coefficients of this function will be chosen by the insurer, exactly as wages above were selected by the firm.

If the claimant has zero initial wealth, his final utility is

$$U\left[-x + w(y) - \xi(x-y)^2\right] = U\left[-x + a + by - \xi(x-y)^2\right]$$

If he maximizes his utility by choosing (over the real line) the claim, he solves

$$dU/dy = 0$$

$$U'\left(-x + a + by - \xi(x-y)^2\right)\left(b + 2\xi(x-y)\right) = 0$$

which means a claim equal to

$$y = x + b/2\xi \qquad (21)$$

This level of claim satisfies the incentive compatibility constraint of the agent. It satisfies also the participation one, since non-participation means filing 0, and we are maximizing over all the possible values of y, 0 included. We will therefore use the incentive compatibility constraint only.

The rational insurer recognizes that the claimant's incentives to engage in claims inflation depend on the indemnity function $w(y)$, i.e., on a and b. A profile w that reduces the incentive of inflating claims would be profitable for the insurer. At the same time, however, if the insurer underpays some claims and creates litigation, it will cost him. The function which relates costs to the difference between the claim y and the indemnity w,

$$c\,(y - w)$$

is assumed to be increasing (with $c\prime > 0$), convex (with $c" > 0$), and equal to zero when the claim is paid exactly ($c(0) = 0$) and when the indemnity is higher than the claim: $w > y$.

The principal reimburses $w(y)$ and suffers the costs $c(y - x)$. Since he is risk-neutral, his expected utility, where the expectation E is taken wrt x, is

$$\mathrm{E}\left[w(y) + c(y - w(y))\right] = \mathrm{E}\left[a + by + c((1-b)y - a)\right]$$

To devise an optimal reimbursement policy, one can write down the insurer problem, as we did in the previous section, as

$$\mathcal{P}_p = \begin{cases} \min_{a,b} \mathrm{E}\left[a + by + c((1-b)y - a)\right] \\ s.t. \\ y = x + b/2\xi \end{cases}$$

Since this particular problem has only one equality constraint, we can substitute it into the objective function and further simplify the problem to an unconstrained minimization one:

$$\mathcal{P}_p = \left\{ \min_{a,b} \mathrm{E}\left[a + b\left(x + b/2\xi\right) + c((1-b)\left(x + b/2\xi\right) - a)\right] \right. \tag{22}$$

Now, let us recall that costs are not zero if and only if the indemnity is lower than the claim:

$$c(y-w) := \begin{cases} 0 & w \geq y \\ > 0 & w < y \end{cases}$$

The inequality $w < y$ is equivalent to

$$y > a/(1-b)$$

and, substituting from the constraint, to

$$x > a/(1-b) - b/2\xi$$

Let us call the right hand side of this inequality x^*. We have

$$c(y-w) := \begin{cases} 0 & x \leq x^* \\ > 0 & x > x^* \end{cases}$$

It follows that the expected value in (22), written out in full and assuming a density $g(x)$ for the damage, is

$$\int_0^{+\infty} \left(a + bx + b^2/2\xi\right) g(x)dx + \int_{x^*}^{+\infty} c((1-b)\left(x + b/2\xi\right) - a)g(x)dx$$

The foc for its minimum wrt a, b are respectively

$$\int_0^{+\infty} g(x)dx - \int_{x^*}^{+\infty} c'(\circ)g(x)dx = 0$$

$$\int_0^{+\infty} \left(x + b/\xi\right) g(x)dx + \int_{x^*}^{+\infty} c'(\circ)(-x + 1/2\xi - b/\xi)g(x)dx = 0$$

where for brevity $\circ := (1-b)\left(x + b/2\xi\right) - a$.

Considering that $\int_0^{+\infty} g(x)dx = 1$, while $\int_0^{+\infty} xg(x)dx$ is the mean of the damage, and denoting the latter by μ, the foc's can be rewritten as

$$\int_{x^*}^{+\infty} c'(\circ)g(x)dx = 1$$

$$-\int_{x^*}^{+\infty} c'(\circ)(-x+1/2\xi-b/\xi)g(x)dx=\mu+b/\xi$$

When claims falsification is prohibitively costly, so that $\xi \to +\infty$, , under some additional assumption on the cost function ($c'(0)=1$), the couple $a = 0, b=1$, which implies $x^*=0$, solves the foc's: indeed, the first becomes

$$\int_{0}^{+\infty} 1g(x)dx=1$$

while the second becomes

$$-\int_{0}^{+\infty} (-x)g(x)dx=\mu$$

which are satisfied by definition.

With infinitely costly claim falsification, it is therefore optimal to pay $w(y) = 0+1y$, i.e., the filed claim y.

Starting with this extreme case, Crocker and Tennyson also show that the optimal choice of b is increasing in ξ. The cost-minimizing indemnity schedule is increasing in falsification costs, so that underpayment should be smaller when it is more costly to falsify the loss. This is the key testable implication of the theory (underpayment should be the greater, the easier it is to falsify the loss): the Authors perform an empirical analysis of insurance settlements for bodily injury in automobile accidents. The analysis confirms their prediction.

An interesting model of the same type, devoted to comparing the incentive effects of no-fault[8] automobile insurance against the tort system, is Cummins et alii (2001).

5.2 LEGAL FEE RESTRICTIONS, MORAL HAZARD, AND ATTORNEY RENTS (R. Santore and A.D. Viard)

The typical principal-agent situation in the legal arena is the relationship between a client (the principal) and his attorney (the agent). Santore and Viard (2001) show that, with both a risk neutral attorney and client, it is efficient for the former to purchase the rights to his client's legal claim by requiring no fixed fee and having zero profits. This is simply an application of the principal-agent theory, with both a risk-neutral agent and principal, which we mentioned above.

In order to appreciate the result, consider a risk neutral client faced with risk-neutral attorneys in a competitive market for legal services.

The agent-attorney's (monetized) effort is E, he faces fixed costs M.

The principal-client expected award (gross of legal fees) is $A(E)$, and it is increasing and concave in the agent's effort, equal to zero for zero effort: $A'(\cdot) > 0$, $A''(\cdot) < 0, A(0) = 0$. We assume that the client and the attorney both know the $A(E)$ function but that the client cannot observe the attorney's effort E. There is no external or business condition uncertainty, X.

[8] No fault insurance has compulsory first-party insurance for personal injuries and restrictions on the ability to sue.

Let F denote the fixed fee and f denote the contingent fee (the fraction of the award received by the attorney), $f \in [0,1]$: the attorney receives $fA(E)+F$, with f and F chosen by the client. The attorney's profits are $fA(E)+F-E-M$. The client's surplus (award net of legal fees) is $(1-f)\,A(E)-F$.

Since the attorney is risk-neutral, he chooses the effort E^* which maximizes his profits (incentive compatibility constraint):

$$fA'(E^*) = 1 \tag{23}$$

The client chooses the fee so as to maximize his own surplus, provided that the attorney will accept (participation constraint) and that he has the incentive to maximize his own profit (incentive compatibility constraint). The former constraint in this context refers to non negative profits for the attorney, $fA(E^*)+F-E^*-M \geq 0$, the second that he exercises the effort E^*, which solves (23). In turn, if the effort exerted is E^* and there is no external uncertainty, there is no uncertainty on the client profit, which is $(1-f)\,A(E^*)-F$. The principal problem therefore is:

$$\mathcal{P}_p = \begin{cases} \max_{f,F}\,(1-f)\,A(E^*)-F \\ s.t. \\ fA(E^*)+F-E^*-M \geq 0 \\ fA'(E^*) = 1 \\ f \in [0,1] \end{cases}$$

It is rather evident that

$$\begin{aligned} f &= 1, \\ F &= -A(E^*)+E^*+M, \\ E^*\text{such that } A'(E^*) &= 1 \end{aligned}$$

solve this problem: there is a 100% contingent fee ($f=1$), while F, the fixed fee, is set at the minimum level which guarantees participation. The latter is asumed to be negative. This solution is intuitive indeed, since, according to (23), the fixed fee does not affect the effort of the attorney.

As opposed to the case in which the agent is risk-averse, the principal does not bear all the risk but obliges the attorney to internalize the results of his effort by imposing the maximum contingent fee.

The attorney receives $A(E^*)+F = E^*+M$, so that he is only compensated for effort and fixed costs (no profit), while the client receives $-F = A(E^*)-E^*-M$, the whole award.

This fee structure gives also the effort which maximizes the combined value of the lawsuit, the sum of the profit of the attorney and surplus of the client, V:

$$V \equiv fA(E)+F-E-M$$

The effort level that maximizes V is the one that satisfies the condition

$$fA'(E) = 1 \tag{24}$$

which in turn coincides with the incentive constraint, if $f = 1$. The client "sells" the lawsuit to the attorney for a price equal to its maximum combined value.

However, the American Bar Association Model Rules of Professional Conduct prohibit such an arrangement. Considering the attorney fee as the sum of a fixed and a contingent part, the Authors show that the ABA restriction is formally equivalent to requiring a minimum fixed fee of zero. It leaves ample room for economic rents, even if the attorneys compete through the contingent fee. The contingent fee is not bid down to the zero-profit level. The restrictions are therefore anticompetitive devices.

References

[1] D. Baron (1986): "Design of Regulatory Mechanisms and Institutions", in D. Schmalensee and R. Willig, *Handbook of Industrial Organization,* North-Holland, Amsterdam

[2] K. J. Crocker and S. Tennyson (2002): Insurance fraud and optimal claim setlement strategies, *The journal of Law & Economics*, vol XLV (2), pp.469-508

[3] J.D. Cummins, R.D. Phillips and M.A. Weiss (2001): "The incentive of no-fault automobile insurance", *The Journal of Law & Economics*, vol XLIV (2), pp.427-464

[4] M. O. Finkelstein (1978): *Quantitative Methods in Law*, The Free Press (MacMillan), New York.

[5] Kreps D. M. (1990): *A course in Microeconomic Theory*, Harvester Wheatsheaf, New York.

[6] P. Jorion (1996): *Value at Risk*, Mc Graw Hill, New York.

[7] Peccati L. (2002): *Metodi quantitativi per giuristi*, Egea, Milano

[8] A. Raviv (1979): "The design of an optimal insurance policy", *American Economic Review*, vol 68 n. 7, pp. 64-79

[9] R. Santore and A.D. Viard (2001): "Legal fee restrictions, moral hazard, and attorney rents", *The Journal of Law & Economics*, vol XLIV (2), pp.549-572

[10] Tirole J. (1998): *The theory of Industrial Organization*, The MIT Press, Cambridge, MA.