Experience Rating as a Mechanism Against Insurance Fraud*†

Abstract

Traditionally, insurance companies attempt to reduce (or even eliminate) fraud via audit strategies, under which claims may be investigated at some cost to the insurer, with a penalty imposed upon insureds who are found to report claims fraudulently. However, it is also clear that, in a multi-period setting, experience rating (increases in subsequent premiums whenever a claim is presented) also provides an incentive against fraud. In this paper, we consider a model in which the *only* mechanism used to combat fraud is experience rating. In this way, our model provides the opposite pole to the pure audit model. We show that there exists an experience rating scheme that will eliminate all fraud in all periods, while guaranteeing non-negative profits to the insurer and participation by the insured.

JEL Classification N^o: D5

Keywords: Insurance fraud, experience rating

Ignacio Moreno, Francisco J. Vázquez and Richard Watt
Dept. Análisis Económico, Fac. C.C.E.E.,
Universidad Autónoma de Madrid,
Ctra. Colmenar km. 15, Madrid 28049, SPAIN

*VERY PRELIMINARY FIRST DRAFT - PLEASE DO NOT QUOTE

†This work is loosely based on chapter 3 of the Phd thesis of the first author. The authors are grateful for financial aid under research project PB98-0059 from DGES.

1 Introduction

Traditionally, the theory of insurance fraud has always taken the stance that the way in which an insurer can combat fraud is by using an audit scheme under which claim reports may be investigated to discover if they are fraudulent or not (see Picard (2000) for a survey of the economic theory of insurance fraud). In short, the audit mechanism will detect[1] fraud with a given probability, and in this case, some type of (financial) penalty can be levied upon the insured. In this way, an insured who misreports a claim enters into a lottery which can be made to have a lower expected utility than the option of honest reporting. Thus fraud can be eliminated, although at some cost to the insurer (the cost of mantaining the audit mechanism). It is straightforward to see that the following two results are true; firstly, the optimal audit mechanism will, typically, depend upon the insured's utility function, and so is "utility specific" in the sence that a different scheme will be optimal for different utility types. Secondly, the insurer may not be interested in eliminating all fraud, since there exists a strictly positive cost of reducing fraud (the audit cost) - indeed, the optimal audit scheme will equate the marginal cost and the marginal income[2] of fraud.

Clearly, in the pure audit model the costs implied by the existence of the option of committing fraud, compared to a situation in which fraud is not possible, are borne by the insurer. However, in a multi-period setting, a second option exists based on premiums that are experience rated (see Lemaire (1985, 1995) for a full actuarial treatment of experience rating in automobile insurance). If an insured's future premiums are increased whenever a claim is presented, there exists a direct incentive against presenting fraudulent claims. Assuming that no auditing what-so-ever is used, then experience rating implies that a fraudulent claim report will provide current positive surplus (over and above the option of honest reporting) to the insured at the cost of future negative surplus (assuming that he continues with the insurance contract). In principle, it should be possible to design an experience rating scheme that provides sufficient future costs for all fraud

[1] By "detect" we mean that sufficient evidence suggesting that fraud has been committed has been gathered, for an impartial third party (e.g. a law court) to pass a verdict declaring that indeed fraud has been committed.

[2] The marginal income from fraud will be measured in terms of fraudulent claims that will not be paid, plus any financial penalty that can be collected from insureds who are found to report fraudulently.

to be eliminated at no cost to the insurer. Showing that indeed this is possible is the objective of this paper.

The way in which fraud is treated in this paper is very one sided, and can be thought of as the opposite pole to the pure audit model. We assume that no audits at all are carried out, and so the only mechanism used to combat fraud is the experience rating scheme. Naturally, the most logical scenario is that neither polar model is appropriate, and instead a mixture of both mechanisms will be optimal. However, before considering the optimal strategy of the insurer as a mixture of all possibilities, we must firstly establish that the experience rating scheme is indeed feasible, which is the modest objective that we set in this paper.

It is, of course, not clear exactly which of the two polar models is prefered by either party to the contract. To see this, note firstly that if one of the underlying characteristics of each polar model is that the insurer gets zero expected profits (i.e. that the insurance market is perfectly competitive), then she will be indifferent between the two options. However, the position of the insured in this case is unclear. In the pure audit model (assuming that all fraud is eliminated), the premium for coverage must be set sufficiently high in order to recover the costs of maintaining the audit mechanism. On the other hand, the pure experience rating system will offer premium uncertainty at a lower expected value (since there is no audit mechanism to pay for)[3] . If we are interested in modelling a monopoly insurer, the insured will always be indifferent between either system since the optimal monopoly strategy will involve saturating the insured's participation contstraint, but in this case it is not clear which mechanism offers greater expected profits.

We also feel that the present model is much closer in spirit to other models of asymmetric information than is the pure audit model. To see this, note that, once an accident has taken place, only the insured is informed of it's true value. Hence, at this point, a traditional adverse selection problem has arisen, in which the insured must report his type (accident value) to the insurer. In all traditional adverse selection models, the self-selecting mechanism works by placing costs on the "good" type so that the "bad" type does not want to pass himself off as being "good",

[3] Note that, since an experience rating scheme implies that each individual's lottery remains dependent upon his own outcome, it must violate the mutuality principle (see Borch (1962)), and hence cannot be first best efficient. It may, however, be second best efficient.

rather than the principal paying for an investigation as to the true identity of the agents. In the fraud model with experience rating, this will imply that insureds who have really had an accident (the "good" types) will suffer the costs implied by the asymmetric information by increased future premiums.

The paper continues as follows; section 2 presents the basic model together with all of the relevant assumptions. In section 3 we show that a pure experience rating mechanism against fraud exists, and we consider its general properties. Section 4 concludes and offers some general directions for future research.

2 The model and assumptions

We assume an infinite number of identical individuals, with utility function for money of $u(x)$ which is assumed to be strictly increasing and concave. We assume that the individuals have, at the beginning of any given period, an endowed wealth of w, and an endowed binomial lottery defined by a loss of L with probability p. Naturally, we assume $L \leq w$. The lotteries in each period are independent, and the outcomes are private information to the individual. The endowed expected utility of each individual in each period is given by

$$Eu(w, L) = pu(w - L) + (1 - p)u(w) \equiv \overline{u}$$

In any given period, an individual may insure his loss lottery with an insurance company. We shall be interested in both the case when the insurer acts in a perfectly competitive market, as well as the case of a monopoly insurer.

For simplicity, we assume that all insurance contracts have full coverage, that is, in the case of a loss, the indemnity is equal to L. Of course, this is equivalent to assuming that, at all times, the premium for an indemnity of x in the case of an accident is given by $px + k$ where k is a constant, in which case, it is well known (and very easy to prove) that all risk averse individuals will always purchase full coverage, so long as k is not greater than the risk premium corresponding to the endowed situation. It is also true that for this type of premium, full coverage is both efficient and profit maximizing so long as the insurer is risk neutral.

While the intertemporal nature of the model demands that we assume a discount factor that converts future utility into present value terms, we shall assume that, in any given period i, the utility of period $i+1$ carries a discount factor of 1, while all periods beyond $i+1$ have a discount factor of 0. In words, the assumption is that, when considering any present choices, the agent considers present utility and follow-on period utility to be equally valuable, and does not weigh any further utilities at all. Aside from the obvious mathematical simplifications that this assumption provides, it can also be thought of as a rudimentary manner in which bounded rationality can be brought into the model. Naturally, since we are considering experience rated contracts with a binomial risk in each period, starting off at period i, in period $i+1$ there are two possible options to take into account (the situation following a claim, and that following a claim free period), therefore, in period $i+2$ there are 4 options, and so on[4] . Surely taking into account all options is far beyond the ability of any reasonable individual. On the other hand, taking present period and next period utility at equal value is also a reasonable approximation in an infinite period world.

This assumption on discounting is not costless. Recall that the incentive not to commit fraud is the future costs to the individual due to the higher premium track that subsequently occurs. On one hand, by keeping the next period discount rate at 1, we are effectively imposing the maximum penalty that our restricted model allows, but on the other hand, we are eliminating from consideration the future steam of utilities that the individual can obtain by continuing with the contract. Whenever the expected premiums in future periods are fair, then these future periods all offer expected utility that is greater than the alternative (the reservation utility). Thus, the short time horizon that we assume in the model comes at the cost of having to make some sort of assumption regarding comittment to continue by the insured. Extending the time horizon that the insured takes into account would have the effect of allowing us to relax any committment assumptions on behalf of the client (assuming that the discount rate is sufficiently high) but would also complicate the calculations required enormously[5] .

[4] In general, when a decision in period i is considered, there are 2^n possible situations in period $i+n$ for $n = 1, 2, ...\infty$.

[5] In any case, there are examples of insurance contracts where the insured does indeed face a legal committment to continue. For example, in order to circulate with a car in most countries, a minimum of insurance coverage (that is typically subject to experience rating) must be purchased to cover third party damages.

We take the following assumption, which seems to be the most reasonable. At the beginning of any given period i the client pays the premium corresponding to the current period and then decides whether or not to sign the renewal contract for the follow-on period. This renewal contract will stipulate two possible premiums, one that corresponds to the contingency that no claim is presented in the current period (period i), and another that corresponds to the contingency that a claim is presented in the current period, and signing it commits the client to paying the premium that corresponds at the beginning of the follow-on period. Effectively then, our assumption is that there is a one period committment on behalf of the client, or that in order to break the contract, the client must inform the insured of this decision one period in advance, as is often the case in real-world contracts[6] .

Since the loss lotteries are discrete (either a loss of L occurs, otherwize the loss is 0), and since only the individual observes the outcome of the loss lottery (that is, whether a loss occurred or not), an opportunity for fraud exists whenever the outcome is that no loss occurs, in which case the individual may fraudulently report a loss to the insurer. We assume that the insurer never audits any claims, but rather simply pays out on any claim that is reported. However, in the period following a claim, the insurer will discriminate the premium demanded of the insured according to whether or not a claim was reported. As an endogenous result in our model it turns out that reporting a claim will result in a higher premium in the follow-on period than not reporting a claim[7] . Hence, reporting a fraudulent claim results in a present benefit to the insured at the cost of a higher future premium. The insurer's objective is to design the premiums in period $i+1$ such that fraud will not be committed in period i, for all i. We assume that in the case of perfectly competitive environments, while the true outcomes of the loss lotteries are private information for the insured, his claims history is common information over all insurers, so that there is no benefit to be gained by leaving one insurer for another.

Naturally, individuals that really suffer an accident will suffer the consequences of the expe-

[6] An alternative to assuming committment is to assume that at the outset a bond is posted which is held by the insurer against any unpaid premium (that is, to cover for any precommitted contract that is reneged upon).

[7] Exactly how each of the two possible follow-on premiums compare with the current period premium will depend on the parameters of the model.

rience rating scheme as if they had acted fraudulently. Given this, the experience rating scheme must be designed under several reasonable conditions. Firstly, it must be incentive compatible (which we take as meaning that it must provide sufficient incentive that fraud is eliminated in all periods[8]). Secondly, it must ensure participation by the insured in the sense that he would rather enter the contractual relationship than to not insure (thereby retaining his endowed expected utility in all periods). Thirdly, the contract must not be such that if the individual legitimately suffers an accident he will still rather present a claim than not (we refer to this as type-2 fraud). Finally, of course, the scheme must offer non-negative expected profits to the insurer, in order that she also is willing to participate.

3 A mathematical formulation of the problem

In this section, we formulate the problem facing the insurer according to a constrained optimization framework. In particular, each of the constraints will be discussed in greater detail in order that the feasible set for all problems can be found. In all cases, we assume that the current period premium is q_0, and in the following period the premium will be $q_1(q_0)$ if a claim is reported in the current period, and $q_2(q_0)$ otherwize. In this way, whether or not a claim is reported in the current period, the situation in the next period is formally identical, but where the new q_0 will be either q_1 or q_2. Given this, we shall concentrate on solving for any particular general period, and then clearly the analysis can always be replicated in the next period, and so on.

3.1 The insured's participation constraint

We take the insured's participation constraint to require that the individual will agree to commit himself to continue the contractual relationship by signing the contract offered for coverage in the follow-on period, which (given that we are assuming that follow-on period utility is not discounted) simply requires that

$$P(q_1, q_2) \equiv pu(w - q_1) + (1 - p)u(w - q_2) \geq \overline{u} \tag{1}$$

[8] Really, it is not clear that the insurer really would like to eliminate all fraud, for may reasons. For example, Crocker and Morgan (1998) show that allowing some fraud can be of aid in aliviating a traditional adverse selection problem.

Clearly, the right-hand-side is the expected utility that is gained in the follow-on period without insurance, while the left-hand-side is simply the expected utility of the follow-on period renewal (assuming honesty), where the premium q_1 must be paid if a loss is suffered (and is reported) in the current period (this happens with probability p) and q_2 otherwize.

3.2 The insurer's participation constraint

The insurer will participate (offer a contract renewal) if doing so does not result in negative expected profits. Consider the follow-on period premiums. Assuming incentive compatibility (see below), with probability p the clients will have suffered an accident in the current period, and hence will have presented a claim and so will face the higher premium, q_1. In the same way, with probability $1-p$ he will face the lower premium, q_2. Hence, given that the expected value of the indemnity (recall, we are assuming full coverage) to the insurer of any client in any period is just $-pL$, in order for the expected profits that the contract offers in the follow-on period to be non-negative, we require[9]

$$p(q_1 - pL) + (1-p)(q_2 - pL) \geq 0 \quad \Rightarrow \quad pq_1 + (1-p)q_2 \geq pL \tag{2}$$

Finally, in the very first period we have $q_1 = q_2 = q_0 \geq pL$. We define

$$B(q_1, q_2) = pq_1 + (1-p)q_2$$

so that the expected profit condition is simply $B(q_1, q_2) \geq pL$.

3.3 The incentive compatibility constraint

The incentive compatibility constraint ensures that no fraud is committed in the current period. In order to correctly consider the incentive constraint, we must assume that no accident has occurred in the current period, which is the only situation in which fraud is possible. Hence the incentive compatibility constraint requires that the expected utility of not having an accident in the current period and not reporting one is not less than the expected utility of not having an accident and yet reporting one. Note that, given the way in which the problem repeats itself from one period to the next, we only need to ensure that fraud in the current period is eliminated assuming honesty in

[9] Note that we are assuming that no cross subsidization over clients occurs.

the follow-on period. This is because, once the follow-on period arrives, it will itself be the current period of a new problem, and hence the replication of the incentive compatibility constraint will ensure that in that period no fraud will be committed.

Recalling that by assumption we have full coverage, in this case, the exected utility obtained in the current period if the insured does not present a fraudulent claim is simply $u(w - q_0)$. Since no claim is presented, the follow-on period premium will be q_2, yielding a follow-on period expected utility of $u(w - q_2)$, where once again, we are assuming honesty in the follow-on period. On the other hand, if (once no accident has occured) the individual presents a claim in the current period, he will get current period utility of $u(w - q_0 + L)$ and follow-on period utility of $u(w - q_1)$. Thus, the incentive compatibility constraint can be expressed as

$$u(w - q_0) + u(w - q_2) \geq u(w - q_0 + L) + u(w - q_1)$$

We rewrite this constraint as

$$I_1(q_1, q_2) \equiv [u(w - q_2) - u(w - q_1)] - [u(w - q_0 + L) - u(w - q_0)] \geq 0 \tag{3}$$

3.4 Type-2 incentive compatibility

Recall, type-2 fraud is defined as having suffered a current period accident but not reporting it. Once again, type-2 fraud is only an issue once an accident has occurred in the current period. The required constraint is

$$u(w - q_0) + u(w - q_1) \geq u(w - q_0 - L) + u(w - q_2)$$

The left-hand-side of the constraint measures the current period utility once an accident has occurred and has been reported, plus the follow on expected utility at the high premium. The right-hand-side measures the current period utility of having an accident and not reporting it plus the follow on period utility at the low premium.

We write the relevant constraint as

$$I_2(q_1, q_2) \equiv [u(w - q_2) - u(w - q_1)] - [u(w - q_0) - u(w - q_0 - L)] \leq 0 \tag{4}$$

3.5 The feasible set

Using (1), (2), (3) and (4) we can now go on to identify the feasible set for both a perfectly competitive insurer and a monopolist. While this can be done mathematically, here we shall restrict our attention to the graphical analysis in the plane (q_1, q_2). Begin with the two participation constraints, (1) and (2).

Firstly consider the insurer's participation constraint, (2). By the implicit function theorem, the slope of any particular contour of the function $B(q_1, q_2)$ is given by

$$\left.\frac{dq_2}{dq_1}\right|_B = -\frac{p}{(1-p)}$$

that is, each contour is a decreasing linear function. Furthermore, since the value of B is increased by increases in either q_1 or q_2, the contours that are further from the origin correspond to greater values of expected profits. One particular contour will correspond to the value $B = pL$, that is, the set of points that satisfy the non-negative expected profit condition is the (strictly convex) set of points on and above the contour $B(q_1, q_2) = pL$. This particular contour intersects the 45^ocertainty line at the point $q_1 = pL$, and touches the horizontal axis at the point $q_1 = L$, and the vertical axis at the point $q_2 = \left(\frac{p}{1-p}\right) L$.

Secondly, consider the insured's participation constraint (1). From the implicit function theorem, we get the result that the slope of any particular contour of $P(q_1, q_2)$ is

$$\left.\frac{dq_2}{dq_1}\right|_P = -\frac{pu'(w-q_1)}{(1-p)u'(w-q_2)} < 0$$

Furthermore, since the second derivative of $\left.\frac{dq_2}{dq_1}\right|_P$ is

$$\left(-\frac{p}{(1-p)}\right)\left(\frac{-u''(w-q_1)u'(w-q_2) + u'(w-q_1)u''(w-q_2)\left(\left.\frac{dq_2}{dq_1}\right|_P\right)}{u'(w-q_2)^2}\right) < 0$$

it turns out that the contours of $P(q_1, q_2)$ are negatively sloped and strictly concave. The contours have slope of exactly $-\frac{p}{(1-p)}$ at the point where they cross the line defined by $q_2 = q_1$.

Clearly, since $P(q_1, q_2)$ is decreasing in both of its arguments, the participation constraint (1) is simply the (strictly convex) set of points that lie on and below the contour $P(q_1, q_2) = \overline{u}$. This particular contour touches the horizontal axis at the point $q_1 = L$, that is, exactly the

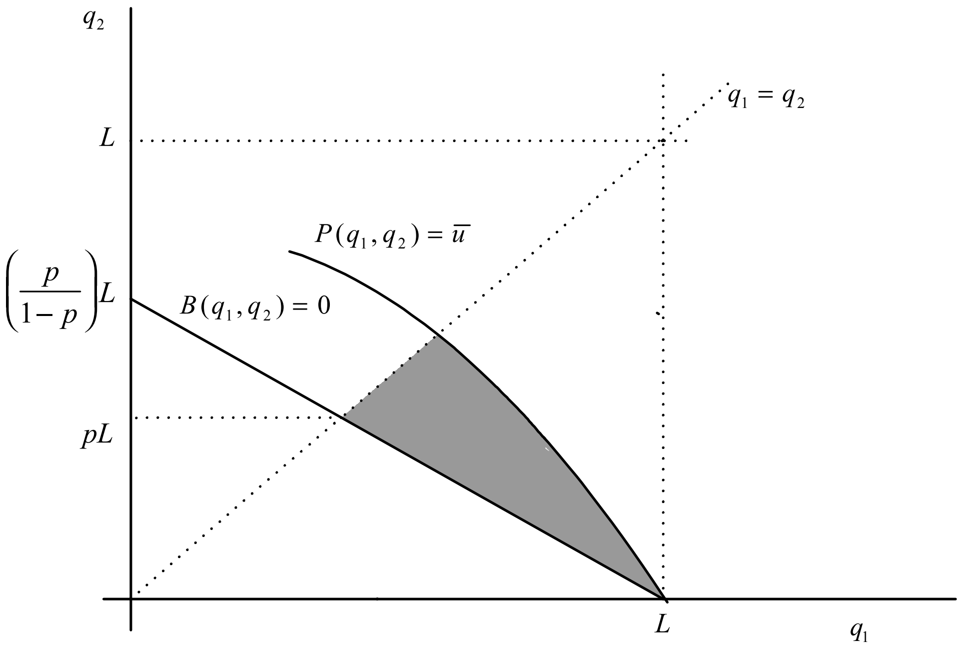

Figure 1:

same point at which the contour $B(q_1, q_2) = pL$ touches the horizontal axis, and it touches the line $q_2 = q_1$ at the point $q_1 = q_{ce}$, where $w - q_{ce}$ is the individual's certainty equivalent wealth. Since the individual is strictly risk averse, we know that $q_{ce} > pL$. The set of points that satisfy, simultaneously, the two participation constraints is shown in figure 1.

Now, consider the incentive compatibility constraints. Consider firstly the type-2 fraud constraint, (4). This constraint saturates along $I_2(q_1, q_2) = 0$. It can be readilly checked from the implicit function theorem that this contour of the restriction is a strictly upward sloping function. For any given value of q_0, this contour passes through the horizontal (q_1) axis at the point where $q_2 = 0$. The resulting value of q_1 is given by

$$I_2(q_1, 0) = [u(w) - u(w - q_1)] - [u(w - q_0) - u(w - q_0 - L)] = 0 \tag{5}$$

Begin by considering $q_0 = 0$, so that (5) reads $[u(w) - u(w - q_1)] = [u(w) - u(w - L)]$, from which it is clearly true that the relevant value of q_1 is precisely $q_1 = L$. But, applying the implicit function theorem to (5), it can be seen that

$$\left.\frac{dq_1}{dq_0}\right|_{I_2(q_1,0)=0} = \frac{u'(w - q_0 - L) - u'(w - q_0)}{u'(w - q_1)} > 0$$

That is, the contour moves to the right as q_0 increases.

Now, condition (4) identifies all the points on or above the particular contour satisfying (5). However, since the limit contour has positive slope, it holds that the entire set of points shown in figure 1 must satisfy the constraint (4) for all $q_0 \geq 0$. Given this, the type-2 constraint can always be ignored.

Finally, consider the incentive compatibility constraint (3). Note directly that, since $u(w - q_0 + L) - u(w - q_0) > 0$, the incentive compatibility constraint can only be satisfied by a point $q_1 > q_2$. Consider the contours (level sets) of the function $I_1(q_1, q_2)$. Directly from the implicit function theorem, the slope of any given contour for any given value of q_0 is simply

$$\left.\frac{dq_2}{dq_1}\right|_{I_1, q_0} = \frac{u'(w - q_1)}{u'(w - q_2)}$$

which is strictly greater than 1 for all feasible points ($q_1 > q_2$). Therefore, a contour that lies (inititally) below the diagonal line $q_1 = q_2$ must approach the line but never reach it, i.e. the contour must lie everywhere below the 45^oline. The result is that each particular such contour must be a positively sloped function that asymptotically approaches the line $q_1 = q_2$. It is important to note that the contours may not be concave functions (their concavity depends on the sign of the third derivative of the utility function).

One particular contour will be that corresponding to equation (3) saturating. Hence, the incentive compatibility condition is simply the set of points on and below this particular contour. To see this, note that starting from a point on the contour, the value of $I_1(q_1, q_2)$ is increased by either an increase in q_1 or a decrease in q_2. This fact also implies that, contours corresponding to higher values of q_0 are located below and to the left of contours corresponding to lower values. To see this, note that a marginal increase in q_0 will change $I_1(q_1, q_2)$ by $u'(w - q_0 + L) - u'(w - q_0)$ which is strictly negative since the utility function is strictly concave.

Finally, consider the point at which the contour $I_1(q_1, q_2)$ passes through the horizontal (q_1) axis.

Lemma 1 *Given $0 \leq q_0 \leq L$, the incentive compatibility contour $I_1(q_1, q_2) = 0$ passes through the horizontal axis at a point $\widetilde{q}_1$ that satisfies $q_0 \leq \widetilde{q}_1 \leq L$, with strict equality holding in both inequalities only when $q_0 = L$.*

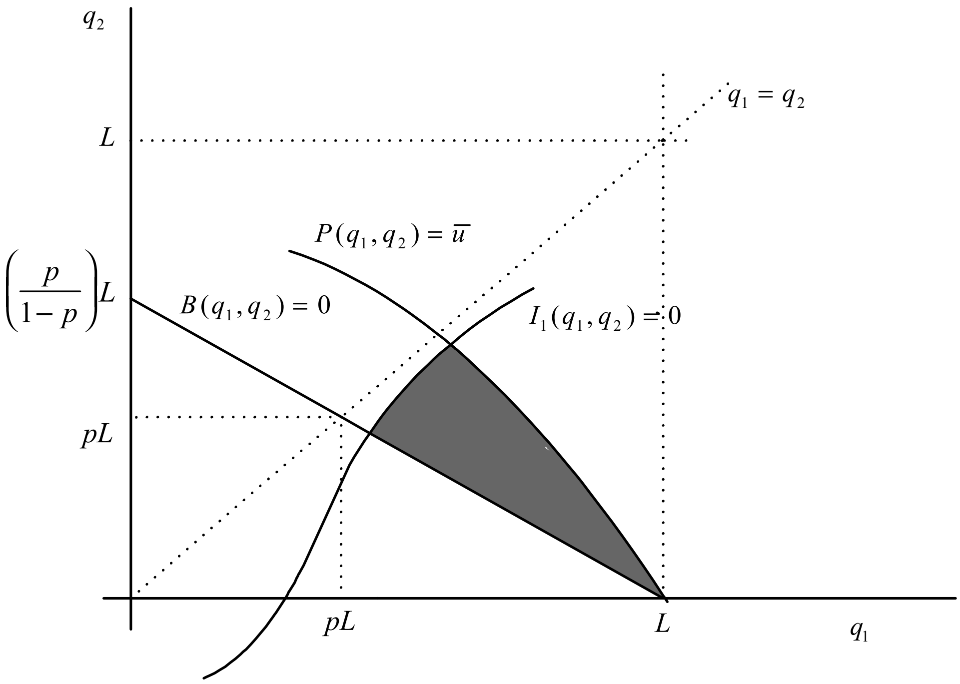

Figure 2:

Proof. Since the required value of q_1 occurs where $q_2 = 0$, it must satisfy

$$[u(w) - u(w - \widetilde{q}_1)] - [u(w - q_0 + L) - u(w - q_0)] = 0$$

that is

$$u(w - \widetilde{q}_1) = u(w) - u(w - q_0 + L) + u(w - q_0) \tag{6}$$

Clearly, so long as $q_0 < L$ we get $u(w) - u(w - q_0 + L) < 0$, so that (6) requires $u(w - \widetilde{q}_1) < u(w - q_0)$, that is, $\widetilde{q}_1 > q_0$. On the other hand, if $q_0 = L$ we have $\widetilde{q}_1 = q_0$. But since the right-hand side of (6) is strictly decreasing in q_0, it must also be true that $\widetilde{q}_1 < L$ for any $q_0 < L$. ■

Given all of the above analysis, we can now draw the feasible set for the problem. This is done for a value of q_0 that is strictly positive but strictly less than L in figure 2. Since $I_1(q_1, q_2) = 0$ is not necessarily a strictly concave function, the upper boundary of the feasible set may not be strictly concave, and so there exists the possibility that the set itself is not convex.

4 The constrained optimization problems

We now go on to consider the two constrained maximization problems, corresponding to a perfectly competitive insurer on the one hand, and a monopolist on the other. We should firstly point out that the (possible) non-convexity of the feasible set requires that, before attempting any constrained optimization, we firstly check the regularity conditions along the frontier of the set, which guarnatees that when we maximize using the Kuhn-Tucker technique, all possible points that are candidate as optimum are considered. This requires that:

- For the restriction $I_1(q_1, q_2) = 0$ we have $\nabla I_1 = (u'(w - q_1), -u'(w - q_2)) \neq (0, 0)$
- For the restriction $P(q_1, q_2) = \overline{u}$ we have $\nabla P = (-pu'(w - q_2), -(1 - p)u'(w - q_2)) \neq (0, 0)$
- For the restriction $B(q_1, q_2) = pL$ we have $\nabla B = (p, 1 - p) \neq (0, 0)$
- At the intersection of the restrictions $I_1(q_1, q_2) = 0$ and $B(q_1, q_2) = pL$ we have ∇I_1 and ∇B linearly independent:

$$\begin{vmatrix} u'(w - q_1) & -u'(w - q_2) \\ p & 1 - p \end{vmatrix} = (1 - p)u'(w - q_1) + pu'(w - q_2) \neq 0$$

- At the intersection of the restrictions $I_1(q_1, q_2) = 0$ and $P(q_1, q_2) = \overline{u}$ we have ∇I_1 and ∇P linearly independent:

$$\begin{aligned} \begin{vmatrix} u'(w - q_1) & -u'(w - q_2) \\ -pu'(w - q_1) & -(1 - p)u'(w - q_2) \end{vmatrix} &= -(1 - p)u'(w - q_1) \cdot u'(w - q_2) - \\ &\quad -pu'(w - q_1) \cdot u'(w - q_2) \\ &= -u'(w - q_1) \cdot u'(w - q_2) \neq 0 \end{aligned}$$

- At the intersection of the restrictions $P(q_1, q_2) = \overline{u}$ and $B(q_1, q_2) = pL$ we have ∇P and ∇B linearly independent:

$$\begin{aligned} \begin{vmatrix} -pu'(w - q_1) & -(1 - p)u'(w - q_2) \\ p & 1 - p \end{vmatrix} &= p(1 - p)[u'(w - q_2) - u'(w - q_1)] = \\ &= p(1 - p)[u'(w) - u'(w - L)] \neq 0 \end{aligned}$$

Given that the regularity conditions are satisfied, we can now go on to consider the two constrained maximization problems formally.

4.1 A perfectly competitive insurer

Assuming that the insurer acts in a perfectly competitive environment, the objective will be to find the two premiums q_1 and q_2 that maximize the insured's expected utility in the follow-on period, subject to the constraints (1), (2) and (3). Thus, the problem can be stated as

$$\max_{q_1,q_2} Eu(q_1,q_2) = pu(w-q_1) + (1-p)u(w-q_2)$$

subject to

$$\begin{aligned} pu(w-q_1) + (1-p)u(w-q_2) &\geq \overline{u} \\ pq_1 + (1-p)q_2 &\geq pL \\ [u(w-q_2) - u(w-q_1)] - [u(w-q_0+L) - u(w-q_0)] &\geq 0 \end{aligned}$$

Clearly, given that the problem is parametized by q_0 (as well as L, w, p and $\overline{u}$) the solution can be expressed as $q^* = (q_1^*(q_0), q_2^*(q_0))$.

We firstly construct the Lagrangian for the problem

$$\begin{aligned} L(q_1,q_2,\lambda_1,\lambda_2,\lambda_3) &= Eu(q_1,q_2) + \lambda_1 I_1(q_1,q_2) + \lambda_2\left[P(q_1,q_2) - \overline{u}\right] + \\ &\quad \lambda_3\left[B(q_1,q_2) - pL\right] \end{aligned}$$

The two first order conditions are

$$\frac{\partial L}{\partial q_1} = \frac{\partial Eu}{\partial q_1} + \lambda_1\frac{\partial I_1}{\partial q_1} + \lambda_2\frac{\partial P}{\partial q_1} + \lambda_3\frac{\partial B}{\partial q_1} = 0$$
$$\frac{\partial L}{\partial q_2} = \frac{\partial Eu}{\partial q_2} + \lambda_1\frac{\partial I_1}{\partial q_2} + \lambda_2\frac{\partial P}{\partial q_2} + \lambda_3\frac{\partial B}{\partial q_2} = 0$$

But since $\frac{\partial P}{\partial q_1} = \frac{\partial Eu}{\partial q_1}$ and $\frac{\partial P}{\partial q_2} = \frac{\partial Eu}{\partial q_2}$ these can be written as

$$\frac{\partial L}{\partial q_1} = (1+\lambda_2)\frac{\partial Eu}{\partial q_1} + \lambda_1\frac{\partial I_1}{\partial q_1} + \lambda_3\frac{\partial B}{\partial q_1} = 0$$
$$\frac{\partial L}{\partial q_2} = (1+\lambda_2)\frac{\partial Eu}{\partial q_2} + \lambda_1\frac{\partial I_1}{\partial q_2} + \lambda_3\frac{\partial B}{\partial q_2} = 0$$

Recall that we must also fulfil the restrictions that

$$\lambda_i \geq 0 \quad i = 1,2,3$$

Now, since the expected utility function Eu is continuous and the feasible set is non-empty, by the Weierstrass Theorem there must exist a global maximum as the solution to the problem. Furthermore, since the regularity conditions are satisfied, we know that the global maximum must satisfy the Kuhn-Tucker conditions.

We shall locate the solution in a few simple steps.

1. The insurer's participation constraint must saturate, that is $\lambda_3 \neq 0$.

   To see this, note that if $\lambda_3 = 0$ then the second first order condition would read

$$\frac{\partial L}{\partial q_2} = (1+\lambda_2)\frac{\partial Eu}{\partial q_2} + \lambda_1 \frac{\partial I_1}{\partial q_2} = 0$$

   However, since

$$\frac{\partial Eu}{\partial q_2} = -(1-p)u'(w-q_2) < 0$$

   and

$$\frac{\partial I_1}{\partial q_2} = -u'(w-q_2) < 0$$

   we have $\frac{\partial L}{\partial q_2} < 0$, that is the condition does not hold.

2. The incentive compatibility condition saturates, that is, $\lambda_1 \neq 0$.

   To see this, note that if $\lambda_1 = 0$ from the first first order condition we have

$$\frac{\partial L}{\partial q_1} = (1+\lambda_2)\frac{\partial Eu}{\partial q_1} + \lambda_3 \frac{\partial B}{\partial q_1} = 0$$

   However, using the relevant substitutions, this requires that

$$\frac{\partial L}{\partial q_1} = -(1+\lambda_2)pu'(w-q_1) + \lambda_3 p = 0 \Rightarrow (1+\lambda_2)u'(w-q_1) = \lambda_3$$

   On the other hand, from the second first order condition we have

$$\frac{\partial L}{\partial q_2} = (1+\lambda_2)\frac{\partial Eu}{\partial q_2} + \lambda_3 \frac{\partial B}{\partial q_2} = 0$$

   Once again, making the relevant substitutions this reads

$$\begin{aligned}\frac{\partial L}{\partial q_1} &= -(1+\lambda_2)(1-p)u'(w-q_2) + \lambda_3(1-p) = 0 \\ &\Rightarrow (1+\lambda_2)u'(w-q_2) = \lambda_3\end{aligned}$$

Therefore, we require that

$$(1+\lambda_2)u'(w-q_1) = \lambda_3 = (1+\lambda_2)u'(w-q_2)$$

However, this implies directly that

$$q_1 = q_2$$

which, as was noted above, cannot ever satisfy the incentive compatibility condtion (a necessary condition for the incentive compatibility condition to be satisfied is $q_1 > q_2$).

3. If the incentive compatibility condition saturates, $\lambda_1 \neq 0$, and if the insurer's participation constraint saturates, $\lambda_3 \neq 0$, then the insured's participation constraint must be slack, $\lambda_2 = 0$, since the point of intersection between $I_1(q_1, q_2) = 0$ and $B(q_1, q_2) = pL$ does not saturate $P(q_1, q_2) = \overline{u}$.

Therefore, we can deduce that there is a single point that satisfies the conditions of the solution; the point characterized by the simultaneous solution to:

$$I_1(q_1, q_2) = 0$$

$$B(q_1, q_2) = pL$$

Let us call this point $q^c = (q_1^c, q_2^c)$. Finally, recall that in the (q_1, q_2) plane, the first of these equations is a strictly increasing function and the second is linear and strictly decreasing. Hence there always exists an intersection, and that intersection is necessarily unique. Furthermore, from Lemma 1, so long as $0 \leq q_0 \leq L$, the fact that $I_1(q_1, q_2) = 0$ is an increasing function implies that it must be true that $q_1^c \in (q_0, L)$, and since we know that $q_1^c > q_2^c$, and that $B(q_1^c, q_2^c) = pL$, it must also hold that $q_2^c \in (0, pL)$.

The graphical intuition behind the perfect competition solution is simple. Note that the objective is to move the indifference curve of the client (contours of the function $P(q_1, q_2)$) as close as possible to the origin, without leaving the feasible set. Clearly, this will imply a point on the contour $B(q_1, q_2) = pL$, that is, the insurer's participation constraint will saturate. Recalling that, due to the fact that in the entire feasible set $q_1 > q_2$, all indifference curves will always be steeper than the contour $B(q_1, q_2) = pL$ at the point that the two intersect. Hence, the point that the client most prefers in the feasible set is the intersection of the contours $B(q_1, q_2) = pL$ and $I_1(q_1, q_2) = 0$. Hence the simultaneous solution to these two equations, for any given value of q_0, gives us the solution to the perfect competition case.

We sum the perfect competition case up in the following theorem.

Theorem 1 *For any given value of q_0, the contract $(q_1^c(q_0), q_2^c(q_0))$ that maximizes the expected utility of the insured (the perfect competition solution), while satisfying incentive compatibility, type-2 incentive compatibility and non-negative profits, is the (unique) point that saturates both (3) and (2). In this solution, the participation constraint of the insured is satisfied but does not saturate.*

Corollary 1 *For all values of q_0 such that $0 \leq q_0 \leq L$, the intersection of the two equations $I_1(q_1, q_2) = 0$ and $B(q_1, q_2) = pL$ occurs at coordinates that satisfy $pL > q_2 \geq 0$ and $L \geq q_1 > q_0$.*

Corollary 1 ensures that, independently of the value of the current period contract premium q_0, all renewals will have prices that are never negative nor greater than L, which is a natural aspect to expect from any experience rated contract. Furthermore, recall that since the solution occurs below the diagonal line, it always corresponds to a situation in which $q_1 > q_2$. However, it is curious that the exact relationship between q_0 and q_2 cannot be determined in general.

4.2 A monopoly insurer

Now assume that the insurer is a monopolist. In this case, the problem that must be solved is to maximize the expected profits in the follow-on period, subject to the constraints (1), (2), and (3). Thus, the problem can be stated as

$$\max_{q_1, q_2} pq_1 + (1-p)q_2 - pL$$

subject to

$$\begin{aligned} pu(w - q_1) + (1-p)u(w - q_2) &\geq \overline{u} \\ pq_1 + (1-p)q_2 &\geq pL \\ [u(w - q_2) - u(w - q_1)] - [u(w - q_0 + L) - u(w - q_0)] &\geq 0 \end{aligned}$$

In the same way as previously, we begin by writing the corresponding Lagrange function:

$$L(q_1, q_2, \lambda_1, \lambda_2, \lambda_3) = B(q_1, q_2) - pL + \lambda_1 I_1(q_1, q_2) + \lambda_2[P(q_1, q_2) - \overline{u}] + \lambda_3[B(q_1, q_2) - pL]$$

The two first order conditions are

$$\frac{\partial L}{\partial q_1} = (1+\lambda_3)\frac{\partial E\pi}{\partial q_1} + \lambda_1\frac{\partial I_1}{\partial q_1} + \lambda_2\frac{\partial P}{\partial q_1} = 0$$

$$\frac{\partial L}{\partial q_2} = (1+\lambda_3)\frac{\partial E\pi}{\partial q_2} + \lambda_1\frac{\partial I_1}{\partial q_2} + \lambda_2\frac{\partial P}{\partial q_2} = 0$$

and we must also bear in mind the restrictions

$$\lambda_i \geq 0$$

Since $B(q)$ is continuous, and since the feasible set is compact and non empty, the Weierstrass theorem guarantees that the problem has a global maximum. On the other hand, since the regularity conditions are satisfied, the global maximum must satisfy the Kuhn-Tucker conditions. As previously, we proceed in simple steps.

1. The insured's participation constraint must saturate, $\lambda_2 \neq 0$.

   To see this, note that if $\lambda_2 = 0$ then from the first first order condition we would have

   $$\frac{\partial L}{\partial q_1} = (1+\lambda_3)\frac{\partial B}{\partial q_1} + \lambda_1\frac{\partial I_1}{\partial q_1} = (1+\lambda_3)p + \lambda_1 u'(w-q_1) > 0$$

   Hence, in this case this first order condition is not satisfied.

2. The incentive compatibility condition must saturate, $\lambda_1 \neq 0$.

   To see this, note that if $\lambda_1 = 0$ from the first first order condition we would have

   $$\frac{\partial L}{\partial q_1} = (1+\lambda_3)\frac{\partial B}{\partial q_1} + \lambda_2\frac{\partial P}{\partial q_1} = (1+\lambda_3)p - \lambda_2 p u'(w-q_1) = 0$$

   However, from the second first order condition we would have

   $$\frac{\partial L}{\partial q_2} = (1+\lambda_3)\frac{\partial B}{\partial q_2} + \lambda_2\frac{\partial P}{\partial q_2} = (1+\lambda_3)(1-p) - \lambda_2(1-p)u'(w-q_2) = 0$$

   and so it would have to be true that

   $$u'(w-q_1) = \frac{1+\lambda_3}{\lambda_2} = u'(w-q_2)$$

   that is, $q_1 = q_2$ which can never satisfy the incentive compatibility condition.

3. If the incentive compatibility condition is saturated, $\lambda_1 \neq 0$, and if the insured's participatin constraint is satisfied, $\lambda_2 \neq 0$, then the insurer's participation constraint must be slack, $\lambda_3 = 0$, since the intersection point between $I_1(q_1, q_2) = 0$ and $P(q_1, q_2) = \overline{u}$ cannot saturate $B(q_1, q_2) = pL$.

Therefore, there any solution that satisfies the Kuhn-Tucker conditions is characterized by the simultaneous solution to the two equations

$$I_1(q_1, q_2) = 0$$

$$P(q_1, q_2) = \overline{u}$$

Let us denote any such point by $q^m = (q_1^m, q_2^m)$. However, as we have already noted, the contour implied by the equation $I_1(q_1, q_2) = 0$ is strictly increasing, and the contour implied by the equation $P(q_1, q_2) = \overline{u}$ is strictly decreasing, and so they can only have a single intersection. Finally, if we define q^* to be the point at which $P(q_1, q_2) = \overline{u}$ intersects the diagonal (clearly, q^* is the greatest premium that the individual would pay for full insurance in any given single period contract), that is $q^* \leftarrow u(w - q^*) = \overline{u}$, then it must always hold that $q_0 < q_1^m \leq L$, and $0 \leq q_2^m < q^*$.

Once again, the graphical intuition is obvious. Clearly, since the constraints are the same as for the perfect competition case, the feasible set is also the same. Indeed, the only change is that now the solution implies moving to higher and higher contours of the function $B(q_1, q_2)$ within the feasible set. Following a very similar reasoning to the perfect competition case, it is easy to see that the solution occurs at the intersection of the two contours $P(q_1, q_2) = 0$ and $I_1(q_1, q_2) = 0$. This is summed up in the following theorem:

Theorem 2 *For any given value of q_0, the contract $(q_1^m(q_0), q_2^m(q_0))$ that maximizes the expected profit of the insurer (the monopoly solution), while satisfying incentive compatibility, type-2 incentive compatibility and participation by the insured, is the (unique) point that saturates both (3) and (1). In this solution, the expected profits of the insurer are strictly positive.*

Corollary 2 *For all values of q_0 such that $0 \leq q_0 \leq L$, the intersection of the two equations $I_1(q_1, q_2) = 0$ and $P(q_1, q_2) = 0$ occurs at coordinates that satisfy $q^* > q_2 \geq 0$ and $L \geq q_1 > q_0$.*

5 The Dynamics of the Solution

The sequence of contracts that is generated by the solution to either problem is interesting in its own right. Here we shall only mention the perfect competition case, although the monopoly case is, of course, very similar. For example, assume that the individual is fortunate and so does not suffer accidents for many periods in a row. In this case, his very first premium will be pL, and his follow-on period premium will be $q_2^c(pL)$, which from Corollary 1 is strictly less than pL. However, since the value of q_0 that will be used to calculate the premiums in the third period is lower than that which was used in the second period, the contour $I_1(q_1, q_2) = 0$ shifts to the left. Given this, we get an intersection between this contour and the line $B(q_1, q_2) = pL$ that is strictly above that of the previous period. Thus the premiums that are offered for the follow-on period to the second have a higher value of q_2 and a lower value of q_1 than those values corresponding to the previous period. Curiously then, although the individual has (by assumption) not reported a claim, his premium is increased, although in exchange the punishment premium is reduced. If the individual continues the sequence without reporting a claim, so that his permium sequence is the sequence of q_2 values that the contract generates oscilates but converges to a fixed value, that we denote by $\overline{q}$. This fixed value satisfies $pL < \overline{q} < q_2(pL)$.

On the other hand, an individual who is very unfortunate will find that his premium sequence (the sequence of q_1 values) monotonically increases and converges to the value L. In any case, given that the contract does eliminate all fraud, and given that the probability of an accident is certainly (in realistic cases) low, the probability of such an outcome is very low.

Perhaps the most realistic case is an individual who reports a claim very infrequently. For such individuals, the premium sequence will converge to $\overline{q}$ until a claim is reported. Then the premium will jump to the punishment premium $q_1^c(\overline{q})$, and will then proceed to converge to $\overline{q}$, although once again oscilating around this fixed value. The premiums that correspond to the first three renewal periods are indicated in figure 3.

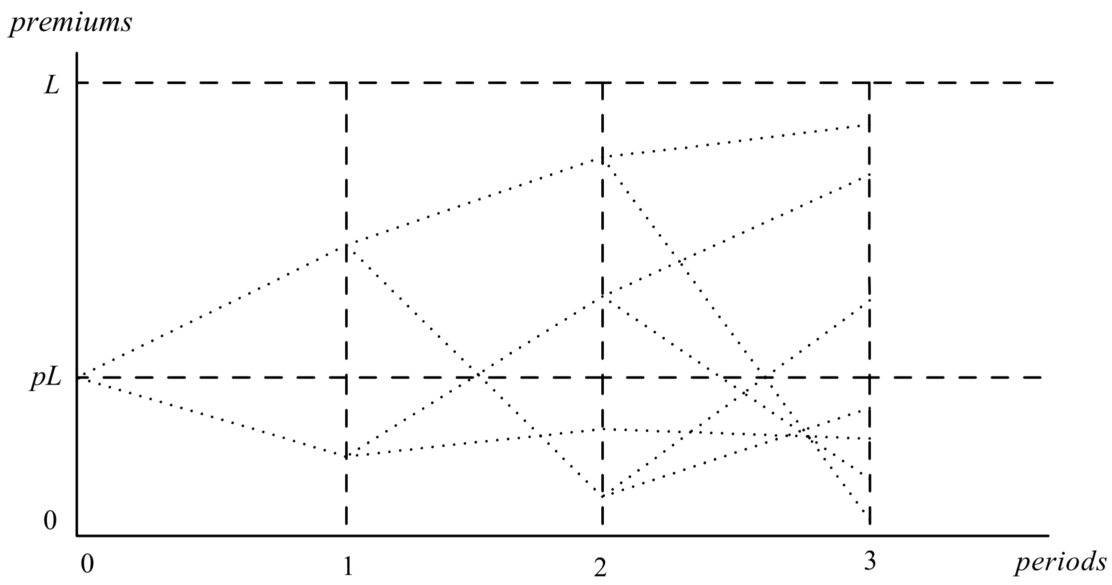

Figure 3:

6 Conclusions

In this paper, we have considered the possibility that an insurer may eliminate all fraud (dishonest claim reporting) using only the incentives that can be provided by experience rated premiums. We have noted that this is, indeed, feasible. In particular, for the cases of both a perfectly competitive insurer and a monopolist, the sequence of premiums that is required to eliminate fraud never implies a negative premium nor a premium that exceeds the value of the insured loss.

Clearly, our model provides an opposite pole to the more traditional case in which fraud is controlled by audits, under which dishonest reporting is punished financially. However, we consider that our model is much closer in spirit to traditional adverse selection models in which the principal does not dedicate resources to evaluating agent type, but rather provides an incentive mechanism under which agents self-select. This is exactly what happens in our model, where agent type is defined by having had an accident or not. Furthermore, our model is also similar to traditional adverse selection models in that the cost of asymmetric information is borne by the "good" agent types (in our model, these are the agents who cannot commit fraud since they have legitimately suffered an accident). In any case, our only objective was to show that all fraud can be eliminated by experience rating, with no audit mechanism at all, and not to advance the conjecture that the model of this paper is in any way more optimal or efficient than a pure audit model. Certainly the optimal anti-fraud strategy would imply some sort of combination of the two options.

Our model can certainly be imporoved upon in many dimensions. The most obvious would be to reconsider the assumption concerning intertemporal discount factors. Amoung other details, the assumption made in this paper has also required us to assume one period committment for the insured, that is, that in order to break a contract sequence, the insured must give notice one period in advance. However, if the insured is given a greater time horizon, calculation costs are increased markedly, and it is doubtful that the marginal benefit in terms of a more consistent model would be greater than the marginal costs in terms of mathematical complications.

Other extentions to the model include (but are not restricted to) extending the dimension of the accident set beyond the two accident states that we have assumed, considering the Pareto comparison of the pure experience rating contract with the pure audit model, and studying the optimal anti-fraud strategy as a combination of the two extremes.

References

Picard, P. (2000), "Economic Analysis of Insurance Fraud", in Dionne, G. (ed.), *Handbook of the Economics of Insurance*, Kluwer Academic Publishers, Boston MA.

Crocker, K. and J. Morgan (1998), "Is Honesty the Best Policy? Curtailing Insurance Fraud Through Optimal Incentive Contracts", *Journal of Political Economy*, 106, 355-375.

Lemaire (1985), *Automobile Insurance: Actuarial Models*, Kluwer Academic Publishers, Boston MA.

Lemaire (1995), *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers, Boston MA.