Do insurance defrauders want to be punished?[1]

D. Alary[2] M. Besfamille[3] [4]

April 3, 2001

[1]We specially thank C. Gollier and B. Jullien for their advice. We are also grateful towards N. Boccard, B. Caillaud, A. Chassagnon, P.-A. Chiappori, M. Guillén, D.-S. Jeon, J.-J. Laffont, J. Lawarrée, P. Legros, A. Pavan, P. Pestieau, P. Picard, P. Rey, G. Roland, P. Seabright, J. Tirole, T. Vergé and B. Villeneuve for helpful comments. Finally we benefited from participants at the 26th Seminar of the "European Group of Risk and Insurance Economists" (Madrid, 1999), the CORE Summer School "Information in Games, Markets and Organizations" (Université Catholique de Louvain, 2000), the JMA Meeting (University of Laval, 2000) and seminars at ECARES (Université Libre de Bruxelles), GREMAQ (Université de Toulouse 1) and Université de Liège. Any remaining errors are of course our sole responsability. M. Besfamille gratefully acknowledges financial support from the Communauté française de Belgique, DRS, ARC 98/03-221.

[2]GREMAQ, Université de Toulouse 1 - Sciences Sociales, Manufacture des Tabacs Bât. "F", 21 allée de Brienne, 31000 Toulouse, France.

[3]CREPP, Université de Liège, Bd. du Rectorat, 7 Bât. 31, 4000 Liège, Belgium.

[4]Corresponding author: Tel: [32](4) 366-3112 / Fax: [32](4) 366-3106 / E-mail: M.Besfamille@ulg.ac.be

Abstract

We analyze a Principal-Agent model of an insurer who faces an adverse selection problem. He is unable to observe if his client has a high risk or a low risk of having an accident. At the underwriting of the contract, the insurer requests the client to declare his risk. After that, the former can costly audit the truthfulness of this announcement. If the audit confirms a false declaration, the insurer is legally allowed to punish the defrauder. We characterize the efficient contracts when this punishment is bounded from above by a legal restriction.

Then, we do some comparative statics on the efficient contracts and on the agent's utility. The most important result of this paper concerns the legal limit to a defrauder's punishment. We prove that there exists a unique value of this legal limit that maximizes the expected utility of a high risk type. Facing this particular value of the legal limit to a defrauder's punishment, the insurer will effectively audit a low risk report. We also show that this particular value increases with the probability of facing a high risk policyholder. Therefore, when this probability is sufficiently high, the nullity of the contract is not enough. From the point of view of a potential defrauder, the law should allow harder sanctions. This is an striking result because the nullity of the contract is a common sanction for this kind of fraud in the USA and in some European countries.

JEL classification: D82, G2, K42

Keywords: Insurance, Principal-Agent, Adverse Selection, Fraud, Audit, Penalties.

1 Introduction

In many countries, there is a growing concern about insurance fraud because its consequent costs are very high. Some estimations corroborate this assertion. Fraudulent claims stand for more than 10% of claims paid in Canada. The *Comité Européen des Assurances* (CEA) considers that insurance fraud represents between 5 and 10% of the total amount of indemnities paid in Europe. In the USA, the annual cost of insurance fraud is estimated to be above $80 billions. In response to that, not only insurers but also many different private and public organizations try to deal with this phenomenon. First of all, there are investigative firms, some with very suggestive names such as "Shadow Chasers" or "Sherlock Investigations", that provide surveillance and information services for insurance companies in the USA or in the UK. In other circumstances, an independent agency is created to investigate on behalf of the insurance companies, as it is the case for the French *Agence pour la Lutte contre la Fraude à l'Assurance* (ALFA).[1] Finally there are independent nonprofit organizations of consumers, government agencies and insurers that are dedicated to combating all forms of insurance fraud through public information and advocacy. They can be national bodies, like the (US) *Coalition Against Insurance Fraud* and the *Canadian Coalition Against Insurance Fraud* or even international groups, like the *International Association of Insurance Fraud Agencies*.

Although there are several forms of fraud that may appear at different stages of an insurance contract, we are only interested in a particular type of misbehavior that takes place during a policy negotiation. At the underwriting, the client has to answer usually some questions about his background information. Moreover, he must communicate truthfully some personal data that is pertinent for the final agreement but which is unknown by the insurer. It is true that countries' legislations differ in the kind of questions the insurer has the right to ask and in the type of information the client is obliged to give. Nevertheless, it is also well documented by practitioners that, at this stage, many policyholders often lie and misreport their private information. As they know that insurers decide to cover them or determine their rate upon those declarations, the clients probably make false declarations in order to

[1] ALFA is a private institution, created by the *Fédération Française des Sociétés d'Assurance* in 1985 and financed by participation fees paid by a pool of insurance companies. ALFA conducts investigations of suspicious cases that are submitted by these insurers.

be effectively covered or to pay lower premiums. We can mention two examples of these attitudes: failing to report an accurate medical history when applying for health insurance and, in automobile insurance, declaring that the usual driver is the middle-aged car's owner while, in fact, it will be his young son.

These misrepresentations seem to occur in many branches of the insurance industry. A recent study published by the (US) *Coalition Against Insurance Fraud* (1997) shows that 63% of the interviewed said that application fraud is very common or fairly common. In another survey conducted by the Florida Insurance Research Center (1991), one third of the interviewed declared that telling lies at the underwriting was acceptable in automobile insurance. Although these figures are not an objective measure of how much widespread this particular type of fraud is, they show the potential magnitude of this problem. In fact, for the insurance industry, it is far from being negligeable. The French agency ALFA found that, in 173 fraudulent cases of car accidents, 75 contracts presented at least one false declaration related to the risk. These false declarations were principally about the driver's identity, its antecedents and the car's use (ALFA, 1999). As stressed by an important member of the Association of British Insurers

> "There is a realization that if even a tiny percentage of policyholders are willing to tell lies, or not disclose relevant information when they take out their policy, it does present the industry with a multi-million pound problem"[2]

If insurers can audit these declarations, they may discover if they were fraudulent. In the USA and in some European countries, insurance companies are allowed to punish the guilty policyholders. But this possibility is strongly regulated by law. In France, the Insurance Code sets precisely the penalties: in case of damage, policyholders who defraud intentionally at the underwriting are not covered and the insurer is allowed to keep the premium (Code des Assurances, Article L.113-8). The Californian legislation is, on the one hand, tougher than the French because unintentional concealment of pertinent information also entitles the insurer to rescind the contract (California Insurance Code, Chapter 3, Article 1, Section 331). But on the other hand, it does not specify anything concerning the premium's return. In the State of New York, the "Insurance Frauds Prevention Act" also includes, for

[2] *Lloyd's List Insurance Day*, October 16, 1993; quoted from Picard (1996).

misrepresentations of information at the underwriting of the contract, the possibility of "a civil penalty not to exceed five thousand dollars", in addition to any other specified criminal liability (New York State Consolidated Laws, Chapter 28, Article 4, Section 404). The first objective of this paper is to characterize the efficient insurance contracts when policyholders that have been found misreporting their personal characteristics can be penalized but below a given exogenous level.

By imposing a limit to a defrauder's sanction, the law restraints the set of contracts that can be offered. Thus insurers' profits and policyholders' rents depend implicitly on the legal framework dealing with insurance fraud. But laws come either directly from the choice of individuals in the case of a *referendum* for a proposition of law or indirectly, via a voted representative, from lobbies' activities and voting. Therefore, if insurers (policyholders) can participate in some way in the legal framework's design concerning insurance fraud, they will try to fix a punishment limitation that maximizes their expected profit (utility). In the USA, both at the federal and at the state level, legislation that provides new and increased criminal and civil penalties has been enacted. In 1991, the Bill 3171 was introduced by the House of Representatives to amend the US Code so that fraud against insurance companies will be subject to strong federal criminal and civil penalties. Many states followed that intent. An weekly magazine for insurers reported that a bill criminalizing some insurance rip-offs was under study in Michigan Legislature.[3] Although the insurance industry has been very active in promoting these legal changes, it seems difficult to explain the strong political support for this hardness of the American legislation only by the action of the insurers' lobby. So the second objective of this paper is to show that not only insurers, which would be natural, but also policyholders, specially those that may be tempted to misreport, do have preferences for high levels of punishment to impose to defrauders.

To deal with these two issues in the simplest way, we present a Principal-Agent model of an insurer who faces an adverse selection problem. Our basic setting is similar to the Stiglitz (1977) framework. The insurer is unable to distinguish between two different types of an agent, namely his low or high

[3] *National Underwriter, Property & Casualty/Risk & Benefits Management*, May 15, 1995. It is interesting to observe that, to support the need for passage of such bill, the article cited a study conducted by the Alliance of American Insurers, illustrating how much widespread falsifications of auto insurance applications were. This misbehavior is exactly the kind of insurance fraud that we study in this paper.

probability of suffering a damage after an accident. We extend the Stiglitz's model in two directions. First of all, we allow the insurer to request his client to declare his risk at the underwriting stage. Then we assume that the insurer has the possibility to audit this declaration. Like many other contributions that have dealt with costly-state verification (for example, Townsend (1979), Baron and Besanko (1984), Reinganum and Wilde (1985), Border and Sobel (1987) and Mookherjee and Png (1989)), we assume that audit is costly but perfect because, if it is performed, it reveals the type of the agent. When the insurer is allowed by law, he can penalize the policyholder that has misreported. But the punishment can not exceed a given exogenous legal limit. The fact that the agent's type can be observable after an audit and the result is verifiable allows the insurer to contract upon the policyholder's declaration and the observation of its truthfulness. As we assume that the insurer has perfect commitment, he includes in the contract the probability of audit and the penalty to impose to a defrauder.

Many articles have studied the impact of misreports in insurance markets. But models with announcements of types at the underwriting stage combined with investigation and verification of these announcements have not been applied to the analysis of insurance fraud.[4] To our knowledge, the only exceptions are the contributions by Doherty and Jung (1993) and Dixit (2000). These papers examine the equilibrium in competitive insurance markets with adverse selection on the risk of damage. Doherty and Jung assume that each agent's type characterizes the support and the average of the distribution of potential losses. Dixit adopts the Rothschild and Stiglitz (1976) framework, where agents differ in their probabilities of suffering an accident.

Although Doherty and Jung and Dixit treat false declarations of risk at the application stage as an adverse selection problem, we prefer to refer to this kind of misbehavior as "fraud". The reason for this qualification is that our model combines the following realistic features: the existence of a common language between the insurer and the policyholder so they can communicate, the legal possibility for the insurer to discriminate between different types by asking the agent to declare his private information, the legal obligation of truthfulness for this declaration (or, equivalently, the qualification of a

[4]Most of the articles that analyze insurance fraud adopt an *ex-post* approach: policyholders either fill in claims for inexistent accidents or commit buildup by inflating the amount of the damage associated with a valid claim. See among others Picard (1996), Bond and Crocker (1997), Crocker and Morgan (1998), Fagart and Picard (1999) and Picard (1999).

misreport as a civil fault or a crime) and finally the existence of an audit technology that enables the insurer to verify the report. When legal restrictions prohibit insurers from using objective but non-observable variables to classify risks, they are not allowed to qualify a policyholder that omits to declare such variable as a "defrauder".

We find that efficient contracts have the same coverage properties than the Stiglitz's equilibrium. A high risk agent receives a full insurance contract whereas a low risk policyholder is partially insured. Although under the Stiglitz's framework a separating equilibrium always exists, low risk types can be excluded from the market if their proportion is relatively low. This is not the case in our model. As the insurer has the possibility to audit and punish a defrauder, he is always able to cover both types of agents. Following the well-known Becker's (1968) argument, the insurer will penalize a defrauder at the maximum legal level. Concerning his audit strategy, our result goes in the same direction than the intuitions that appear in Border and Sobel and in Mookherjee and Png, although their setting is different to ours. As only the high risk agent has incentives to mimic the low risk policyholder, only a low risk report will be audited. Finally, we show that different cases of efficient contracts arise. They depend crucially on the legal limit to punishments.

In the second part of the article, we do some comparative statics on the efficient contracts and on the agent's rent. Our most important result is the existence of a unique level of the legal limit for a defrauder's punishment that maximizes the high risk type's expected utility. This particular level will imply afterwards a punishment that will be neither negligeable nor infinite. This result seems striking because one could expect that what defrauders want is precisely not to be punished. But if potential defrauders have the power to design a legal framework that avoids punishments, the insurer will contractually react. He will not audit and, moreover, he will distort too much the contract offered to a low risk client, which is at the detrimental of the other type. This level of the legal limit for a defrauder's punishment is increasing in the probability that the agent is a high risk one. When this probability is sufficiently high, the nullity of the contract is not enough. From the point of view of a potential defrauder, the law should allow harder sanctions. This is an striking result because the nullity of the contract is a common sanction for this kind of fraud in the USA and in some European countries. We believe that our results could serve as an explanation for the strong political support for some recent changes in the American legislation

concerning insurance fraud.

The paper is organized as follows. In the next section, we present the model. Then we analyze the benchmark, where no audit is possible. Next, we characterize the different contracts that arise in equilibrium when the insurer audits but the penalties he can impose are bounded from above by law. Then we discuss the existence of an optimal level of punishment that the insurer can impose to defrauders. In the conclusion, we summarize the main results of the paper and we comment about future research on the political economy of anti-fraud legislation in insurance markets. All proofs are shown in the Appendix.

2 The model

The policyholder is a risk-averse agent which has an initial wealth ω. With a strictly positive probability π, he may have an accident. In that case, he suffers a loss $\ell < \omega$. He has a von Neumann - Morgenstern utility function u that verifies the usual properties of monotonicity and strict concavity. Let $u_0^{NA} = u(\omega)$ and $u_0^A = u(\omega - \ell)$ be the agent's utilities in the states of Nature "no accident" (NA) and "accident" (A) respectively.

The agent can be of two different types $j \in J = \{H, L\}$. He can face a high risk (H) or a low risk (L) of having the accident, implying that $\pi_{_H} > \pi_{_L}$. Let $\underline{\mathcal{U}}_j \equiv (1 - \pi_j)u_0^{NA} + \pi_j u_0^A$ be the reservation level of expected utility of an agent of type j. The agent knows privately his type. The probability that the agent is low risk is denoted by $\mu \in (0, 1)$.

The insurer is risk-neutral. He is completely uninformed about the agent's type but knows the probability μ, together with the value of ω and ℓ. At the beginning of the contracting stage, he requests the client to declare his risk. We denote by $\widetilde{j} \in \{L, H\}$ this announcement. This formalization is a reduced form of a more sophisticated model. In such model, the policyholder declares some personal and objective characteristics (like age, domicile or medical record) that are unobservable for the insurer. Upon this declaration, the insurer estimates the risk of loss. Our assumption about the policyholder's knowledge of his probability of accident is equivalent to say that he knows the insurer's technology of estimation of the risk and therefore he is also able to estimate his own risk. In fact, this is not so unrealistic. Many US States Insurance Departments explain to consumers why insurers take into account those mentioned factors to consider applications for coverage. Their

consumer publicly provided guides explicitly describe the reason: insurers' statistical data show correlation between those factors and the probability of having an accident (see, for example, Kansas Insurance Department (2000)).

The insurer observes if the accident has effectively occurred. After that, he can audit the reported type with probability γ.[5] The audit is perfect in the sense that, if it is done, it enables the insurer to discover the agent's type. We assume that the insurance company has an audit sector with fixed capacity and decreasing marginal productivity. Therefore we formalize his expected audit cost as a strictly increasing and convex function $c(\gamma)$ that verifies the following properties

$$c(0) = 0,\ c'(0) = 0 \ \text{ and } \lim_{\gamma \to \overline{\gamma}} c(\gamma) = +\infty$$

where $\overline{\gamma}$ is the highest frequency of audit attainable, given the technology of inspection available to the insurer. Instead of using transfers as control variables (like in most contributions in insurance theory), we adopt a dual approach as Grossman and Hart (1983) did. In order to solve the model in the utility space, we have to define the strictly increasing and convex function $v \equiv u^{-1}(\,)$. In this setting, an insurance contract $\mathcal{C}$ has the following shape

$$\mathcal{C} = \left\{ \ \widetilde{j} \in \{L, H\} \ \ \rightarrow \ \ \left(u^{NA}_{\widetilde{j}}, u^{A}_{\widetilde{j}}, \gamma_{\widetilde{j}}, u^{D}_{\widetilde{j}} \right) \ \right\}$$

After an announcement of the risk, the insurer offers the policy, indicating that he commits to leave to his client a level of utility in each final state and to a probability of audit.[6] We assume total commitment for this contract and we define

- u^{NA}: the utility when the agent accepts the contract and has no accident,
- u^{A}: the utility when the agent takes the contract and has an accident,
- u^{D}: the utility to be left to a defrauder after the audit founds a misreport.

[5] For the sake of simplicity, we rule out the possibility of audit before the accident. Although this option could in principle enrich the analysis of the efficient contract offered by the insurer, it would not change qualitatively the main result of this article, namely the fact that a defrauder wants a punishment if he is caught.

[6] It is easy to show that, in this framework, the insurer never gains by offering to the client a contract where the latter is not obliged to announce his risk.

Institutional constraints: The insurer is prevented from rewarding the policyholder after an audit has confirmed his report. Hence, in that case, the agent's utility is the same than with no audit. Moreover the insurer can not over-insure the client so $u^{NA} \geq u^{A}$.[7] Finally there is also a legal lower-bound $\underline{u}$ on the level of utility to be left to the client in any final state.

3 The efficient contracts

In this section, we analyze the policies offered to the agent. For any given value of $\underline{u}$, the efficient contract maximizes the insurer's expected profit. First of all, to have a benchmark for comparison, we study the situation when the insurer can not audit. Then we look at a more general framework, where we consider this last possibility and we include punishments for defrauders.

3.1 The efficient contracts without audit

As the insurer is unable to audit, we do not have to consider the legal lower-bound $\underline{u}$. This framework corresponds to the Stiglitz (1977) model. The most important results of his model are gathered in the next proposition. Let $\mathcal{U}_j$ be the expected utility of an agent of type j and u_j be the utility when the policyholder gets a full-insurance contract that sets $u_j^{NA} = u_j^A$. From now on, the hats will characterize the Stiglitz's solutions.

Proposition 1 *The insurer offers two different incentive-compatible contracts. The high risk agent always receives a full insurance contract. But the equilibrium depends upon the probability of a low risk agent μ. There exists a threshold $\hat{\mu}$ such that the offered contracts are as follows:*

- $\forall \mu < \hat{\mu}$, $\hat{u}_L^{NA} > u_0^{NA}$, $\hat{u}_L^A < u_0^A$ *and* $\hat{u}_H = \underline{\mathcal{U}}_H$
- $\forall \mu \geq \hat{\mu}$, $u_0^{NA} > \hat{u}_L^{NA} > \hat{u}_L^A > u_0^A$, $\mathcal{U}_L = \underline{\mathcal{U}}_L$ *and* $\hat{u}_H > \underline{\mathcal{U}}_H$

Although a high risk agent has incentives to declare to be a low risk, the insurer never offers pooling contracts because he gains with the discrimination.[8] As the high risk policyholder have the highest willingness to be

[7] These two institutional constraints are introduce to deal with ex-post moral hazard problems, either from the policyholders (setting excessive number of claims to obtain the reward) or from the insurer (declaring he has not seen the accident).

[8] The only pooling equilibrium arises when $\mu \to 1$. In that case, the insurer offers a unique contract, setting $\hat{u}_H = \hat{u}_L^{NA} = \hat{u}_L^A = \underline{\mathcal{U}}_L$.

covered, he always gets a full insurance contract. But, in order to deal with incentive-compatibility, the insurer has to distort the contract offered to a low risk agent. Therefore, as $\hat{u}_{L}^{NA} > \hat{u}_{L}^{A}$, the latter is only partially insured.

When the probability μ is below the threshold $\hat{\mu}$, the complement probability of a high risk individual is relatively large. The insurer has to propose to the low risk agent a contract so distorted that in fact it will not be accepted. So this type of agent is no longer covered against the loss. By doing so, the insurer is able to extract all high risk policyholder's informational rent but at the cost of obtaining positive profits only for one type of agent.

When the probability μ is above the threshold $\hat{\mu}$, the insurer covers both type of agents by discriminating between them. The low risk gets a partial insurance contract that sets him to his reservation level of expected utility $\underline{\mathcal{U}}_{L}$. The other agent receives a strictly positive informational rent $\hat{u}_{H} > \underline{\mathcal{U}}_{H}$. In this case, the final utilities $\hat{u}_{j}$ depend upon the probability of the low risk agent μ. The utilities $\widehat{u}_{L}^{A}$ and $\widehat{u}_{H}$ are increasing functions of μ because, as this probability increases, the insurer can reduce the distortions on the low risk's contract, distortions with respect to the full-insurance contract.

3.2 The efficient contracts with audit and penalties

Now we characterize the contracts offered by the insurer when he audits but also faces the legal lower-bound $\underline{u}$. Instead of solving directly the insurer's optimization problem, we simplify it through a sequence of lemmas and propositions. First of all, we prove an important intermediary result, which states the necessity of penalizing a defrauder, in terms of utility, to audit with strictly positive probability.

Lemma 1 *If an efficient contract sets $\gamma_{\tilde{j}} > 0$ then it must verify $u_{\tilde{j}}^{D} \leq u_{\tilde{j}}^{A}$.*

The insurer does not audit if he has to leave to a defrauder a higher utility than to an agent who has truthfully reported his type. The role of an audit, combined with a punishment, is to reduce the incentives to misreport. But the insurer can not threaten a defrauder with this possibility if for the latter there is no real loss, in terms of utility, between truthful reporting and misreporting. Therefore, in that case, it is not worth for the insurer to audit because he will bear the expected cost. We call the difference in utility $u_{\tilde{j}}^{A} - u_{\tilde{j}}^{D}$ a "punishment". This is a necessary condition to threaten a defrauder with the possibility of an audit and therefore the discovery of his

type. As we want to restrict our analysis to the punishment's side, we focus our attention on pairs of parametric values $(\underline{u}, \mu)$ such that the lower-bound verifies $\underline{u} \leq \hat{u}_{L}^{A}(\mu)$.[9] Next we prove the following result.

Proposition 2 *The insurer prefers contracts such that each type of agent reports truthfully.*

Although our model is theoretically similar to Border and Sobel (1987), we impose more restrictions than them, specially the fact that truthful reports can not be rewarded. So we can not directly apply their modified version of the Revelation Principle, set in the first proposition of their article. Nevertheless we prove that the insurer does not gain by offering contracts that induce a misreport in equilibrium.[10] Therefore, the efficient contracts are incentive-compatible and all their components are contingent on the announcement of the type j. These contracts solve the following program

$$\mathcal{P}\begin{cases} Max \ \underset{j}{\mathbb{E}}\,\Pi & & \\ subject\ to & & \\ \forall\ j & \mathcal{U}_j \geq \underline{\mathcal{U}}_j & IR(j) \\ \forall\ j & \mathcal{U}_j \geq \mathcal{U}_{j\,j'} \quad j' \neq j & IC(j) \\ \forall\ j & u_j^{NA} \geq u_j^{A} & NOI(j) \end{cases}$$

where

$$\begin{aligned} \underset{j}{\mathbb{E}}\,\Pi = \ & w - (\mu\pi_{_L} + (1-\mu)\pi_{_H})\ell \\ & -\mu\left[(1-\pi_{_L})v(u_{_L}^{NA}) + \pi_{_L}(v(u_{_L}^{A}) + c(\gamma_{_L}))\right] \\ & -(1-\mu)\left[(1-\pi_{_H})v(u_{_H}^{NA}) + \pi_{_H}(v(u_{_H}^{A}) + c(\gamma_{_H}))\right] \end{aligned}$$

is the insurer's expected profit,

$$\mathcal{U}_j \equiv \ (1-\pi_j)u_j^{NA} + \ \pi_j u_j^{A}$$

[9] Recall that $\hat{u}_{_L}^{A}$ is the Stiglitz's level of utility of a low risk agent that has suffered an accident but is covered. As we mentioned before, this value depends on the probability of a low risk agent μ.

[10] We could have considered more general mechanisms. For example, after an announcement $\widetilde{j}$, instead of giving a unique contract $\left(u_{\widetilde{j}}^{NA}, u_{\widetilde{j}}^{A}, \gamma_{\widetilde{j}}, u_{\widetilde{j}}^{D}\right)$, the insurer can offer a menu of contracts to the policyholder. But our result also holds in this more general case.

is the expected utility of a type j policyholder that announces truthfully his type and

$$\mathcal{U}_{j'j} \equiv \ (1-\pi_j)u_{j'}^{NA} + \ \pi_j[\gamma_{j'}u_{j'}^{D} + (1-\gamma_{j'})u_{j'}^{A}] \quad j' \neq j$$

is his expected utility if he misreports. Let $IR(j)$, $IC(j)$ and $NOI(j)$ be the participation constraint, the incentive-compatibility constraint and the no-over insurance constraint for each type. A contract that verifies all these constraints is called "feasible". Next we show a result that goes in the sense of Becker's (1968) well-known proposition.

Lemma 2 *If an efficient contract sets a strictly positive probability of audit* $\gamma_j > 0$, *the level of utility to be left to a defrauder* u_j^D *is equal to the minimum legal level* $\underline{u}$.

There is no gain for not penalizing at the maximum a defrauder since, by Proposition 2, this will never happen in equilibrium. Although this model does not give rise to fraud in equilibrium, this is an out-of-equilibrium possibility whose outcome depends crucially on the minimum legal level $\underline{u}$. And in fact it has to be considered by the insurer at the contract design stage. Next we show that, for a high risk agent, his contract does not qualitatively differ from the Stiglitz's contract.

Lemma 3 *A high risk policyholder is always fully insured.*

As a high risk agent has the highest willingness to be fully covered, the insurer maximizes his profits by given him a full-insurance contract. The next lemma states that, in equilibrium, only one type of agent has an incentive to misreport.

Lemma 4 *An efficient contract verifies that the incentive-compatibility constraint* $IC(L)$ *is slack, the incentive-compatibility constraint* $IC(H)$ *binds and a high risk individual is never audited.*

Since the low risk type has no incentives to defraud, the audit probability for an agent that declares to be of a high risk can be set equal to zero. But the opposite obviously does not hold. In Border and Sobel's terminology, a high risk policyholder "is attracted by" a low risk type and thus has incentives to defraud. Therefore it may be optimal to audit the low risk type with a

positive probability, set at the lowest level that is compatible with a high risk individual being indifferent between the two contracts. A similar result was noted by Border and Sobel (1987) and Mookherjee and Png (1989). Applying the previous lemmas, we can also show another result.

Lemma 5 *An efficient contract verifies that the participation constraint $IR(L)$ binds.*

The insurer maximizes his profit by extracting all rents from a low risk individual. This type of agent is indifferent between buying the contract or not. We assume that, if the contract implies a non-negative coverage of the risk, a low risk individual will take it. Finally we can exhibit some conditions under which the no-over insurance constraint $NOI(L)$ is slack.

Lemma 6 *If, in equilibrium, $\gamma_{_L} < \tau \equiv \frac{\triangle\pi}{\pi_{_H}(1-\pi_{_L})}$, the low risk policyholder is partially insured.*

In fact, the insurer faces an incentive problem that can be solved in two different ways. On the one hand, when the technology of audit is relatively inefficient so that the expected cost of audit increases very fast with respect to the probability γ, the insurer uses this option with a low probability. So, to relax the binding incentive-compatibility constraint $IC(H)$, he offers a partial-insurance contract to a low risk agent. This has a second-order negative effect on the insurer's expected profit but the consequent reduction in the high risk utility $u_{_H}$ has a first-order positive effect. On the other hand, when the technology of audit is very efficient, the insurer can combat fraud by auditing more frequently. If this is the case, he does not need to distort too much the contract offered to a low risk policyholder. In fact, it may be optimal to set, for a low risk agent, $u_L^A > u_L^{NA}$. As this is not allowed by institutional constraints, in that case the insurer has to offer a full-insurance contract for this type of client. As most of the contributions to insurance fraud consider that audit has a linear expected cost, they can not show the impact of the marginal cost on the distortions for low risk agents. From now on, we assume that the audit technology is very inefficient and so costly that, at the optimum, $\gamma_{_L} < \tau$. This is possible if the upper bound $\overline{\gamma}$ is below τ. Using the previous results, we can rewrite the initial problem $\mathcal{P}$ as follows:

$$\begin{cases} Max \underset{j}{\mathbb{E}} \Pi & \\ subject\ to & \\ \mathcal{U}_{_L} = \underline{\mathcal{U}}_{_L} & IR(L) \\ u_{_H} \geq \underline{\mathcal{U}}_{_H} & IR(H) \\ u_{_H} = \mathcal{U}_{_{LH}} & IC(H) \\ \gamma_{_L} \geq 0 & \end{cases}$$

The shape of the efficient contracts depend on the different values of the probability μ and the minimum level of utility $\underline{u}$ that the insurer is obliged to leave to a defrauder. Three different types of contracts may arise in equilibrium. They are characterized as follows:

- Case A: $\gamma_{_L} > 0$, $u_0^{NA} > u_{_L}^{NA} > u_{_L}^{A} > u_0^{A}$ and $u_{_H} > \underline{\mathcal{U}}_{_H}$
- Case B: $\gamma_{_L} > 0$, $u_0^{NA} > u_{_L}^{NA} > u_{_L}^{A} > u_0^{A}$ and $u_{_H} = \underline{\mathcal{U}}_{_H}$
- Case C: $\gamma_{_L} = 0$, $u_{_H} \geq \underline{\mathcal{U}}_{_H}$ and $u_{_L}^{NA} \geq u_{_L}^{A}$ (the Stiglitz solution).

In cases A and B the insurer covers both types of agent. But in the second case, he is also able to extract all the surplus of a high risk individual. This last configuration was not achievable in the Stiglitz framework. But here the audit combined with a penalty enables the insurer to relax the binding incentive-compatibility constraint $IC(H)$. Thus the insurer can increase the coverage of a low risk agent. In the next figure we show the parametric regions in the space $(\underline{u}, \mu)$ where these different cases arise.

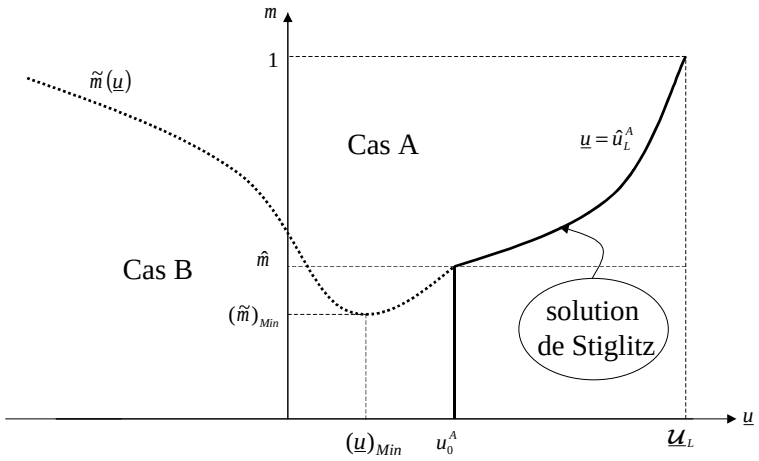


Figure 1: The efficient contracts

In order to get an intuition of the frontiers, recall that we have restricted $\underline{u}$ to verify $\underline{u} \leq \widehat{u}_L^A(\mu)$. When $\underline{u} = \widehat{u}_L^A$, it is not worth for the insurer to audit. Hence he sets the Stiglitz solution, which is drawn as a kinky frontier. First of all, let $\mu > \widehat{\mu}$. Departing from the Stiglitz's case, the insurer audits only when $\underline{u} < \widehat{u}_L^A$. But he has to leave a strictly positive rent to a high risk policyholder. When $\underline{u}$ decreases and attains the frontier denoted by to $\widetilde{\mu}(\underline{u})$, the insurer is able to extract all high risk agent's rent because the punishment for a fraud becomes tougher. When $\underline{u} \leq u_0^A$ and $\mu \leq \widehat{\mu}$, although the insurer can limit or extract a high risk's rent, he is also able to cover a low risk agent. The reason is that, as the penalty is high in that parametric region, the threat of audit relaxes the efficiency-rent trade-off. Finally, when the probability of facing a low risk agent is so low that $\mu < (\widetilde{\mu})_{Min}$, the insurer cannot leave any positive rent to a high risk agent. Although in the left inferior part of the figure the legal lower-bound $\underline{u}$ can be very low, the probability of having to leave such rent is so high that the trade-off goes in the direction to set this type of agent to its reservation level of expected utility $\underline{\mathcal{U}}_H$.

4 The defrauders' most preferred legal limit to punishments

The insurer's expected profit cannot decrease when the legal limit $\underline{u}$ decreases because the set of feasible contracts becomes larger. The common sense would say the opposite for the agent: his expected utility cannot decrease when $\underline{u}$ increases, specially if he is a high risk type prone to defraud. In fact, due to the second-best nature of our model, this is not true. The policyholder has a most preferred value for the legal limit $\underline{u}$ and this particular value is not maximal. As we are only interested on fraud and on the punishment's side of the model, we only consider values of $\underline{u}$ that are below $\widehat{u}_{L}^{A}$. By doing that, we do not want to take into account restrictions to the insurer's behavior if he does not audit. To have explicit results, we adopt the following functional forms:

- $v(u) = \frac{-\ln(1-u)}{\beta}$ where β is the coefficient of absolute risk aversion of the underlying (CARA) utility function u
- $c(\gamma) = c\frac{\gamma^2}{2}$.

A low risk individual always obtains the same level of expected utility because his individual rationality constraint binds. Therefore he is indifferent between any minimal level of utility $\underline{u}$. But this is not the case for a high risk policyholder. Under the Case A and the Stiglitz solution, he obtains strictly positive informational rents that are not *a priori* easy to compare. In fact, we are able to show the following result, which is the most important of this paper.

Proposition 3 *When the probability of facing a low risk agent is above* $(\widetilde{\mu})_{Min}$ *and the derivative of the expected marginal cost exceeds* $\widehat{c}$*, there exists an unique level of minimal utility to leave to a defrauder such that, facing this legal limit,*

- *the insurer audits,*
- *the potential high risk defrauder would be punished if he misreports but his expected utility is nevertheless maximal.*

As we can see in Figure 1, a high risk policyholder may obtain a strictly positive rent only when $\mu \geq (\widetilde{\mu})_{Min}$. When the probability of facing this type of agent is too important because $1-\mu \geq 1-(\widetilde{\mu})_{Min}$, it is optimal for the insurer to leave him with his reservation level of expected utility $\underline{\mathcal{U}}_{H}$. But when this probability $(1-\mu)$ is not to high, the insurer faces a trade-off between efficiency and rent extraction. Although this result about the value $\underline{u}^{*}$ and its implications in terms of punishment for a potential defrauder seems counter-intuitive, its explanation can be seen in a constructive way, departing from a situation where, for a given $\mu > \widehat{\mu}$, the minimal level of utility $\underline{u} = \widehat{u}^{A}_{L}$. We know that, in that case, the insurer does not audit and sets the Stiglitz's solution. Let $\underline{u}$ slightly decrease. Facing the new legal limit, it is optimal for the insurer to start auditing the low risk agent in order to reduce the other type's incentives to fraud. Although audit is costly, it enables the insurer to attenuate the distortion of the Stiglitz's partial insurance contract offered to the low risk policyholder at the cost of an increase in the rent of the high risk agent that is lower than when the audit was not possible. So, starting from the Stiglitz's values $\widehat{u}^{NA}_{L}$ and $\widehat{u}^{A}_{L}$, the insurer decreases u^{NA}_{L} and increases u^{A}_{L}. The efficiency gains are high enough to offset an increase in u_{H}. If $\underline{u}$ decreases more, the efficiency of the audit increases because the loss in utility for a defrauder $(u^{A}_{L} - \underline{u})$ increases, which relaxes more the incentive-compatibility constraint of the high risk type. Thus, the marginal cost of satisfying this constraint $IC(H)$ decreases. Therefore the insurer obtains more efficiency gains and can allow a new increase in u_{H}. But if $\underline{u}$ continues to decrease, it attains a value denoted by $\underline{u}^{*}$ where the efficiency gains are lower than the incentive gains. Therefore the insurer starts to extract the rent from the high risk agent. His expected utility attains a maximum level u^{*}_{H} and then decreases. We call the value $\underline{u}^{*}$ the high risk's optimal level of minimal utility to leave to an audited defrauder.

In order to completely characterize the set of values $\underline{u}^{*}$, we do some comparative statics. We prove the following result, which characterizes the mentioned set.

Proposition 4 *When the probability of facing a high risk policyholder increases, his optimal level of minimal utility to be left to an audited defrauder* $\underline{u}^{*}$ *decreases and the punishment* $(u^{A}_{L} - \underline{u}^{*})$ *increases.*

In the Stiglitz's model, when the proportion of high risk policyholders in the population becomes larger, the insurer faces an increasing problem of potential fraud because more people want to misreport. As a consequence, the

distortions in the contract offered to the low risk agents are more important because the trade-off between efficiency and rent extraction is more favorable to the second goal. In our setting, departing from the Stiglitz's solution, the higher is the probability of facing a high risk agent, the greater are the gains from auditing a low risk policyholder. There are two reasons for that. First of all, when it is more likely to face a high risk agent, there are more efficiency gains on the other type from relaxing the incentive-compatibility constraint $IC(H)$ via an audit. Second, as this audit is done only for a low risk report, *ceteris paribus* it is less frequent and therefore it costs less. Hence the insurer can audit more. Recall that, for a given μ, when $\underline{u}$ departs from $\widehat{u}_{L}^{A}$, the efficiency of the audit increases. Therefore these two effects can reinforce each other the lower $\underline{u}$ is. In the next figure, we show the locus $\underline{u}^{*}(\mu)$ in the $(\underline{u}, \mu)$ space.

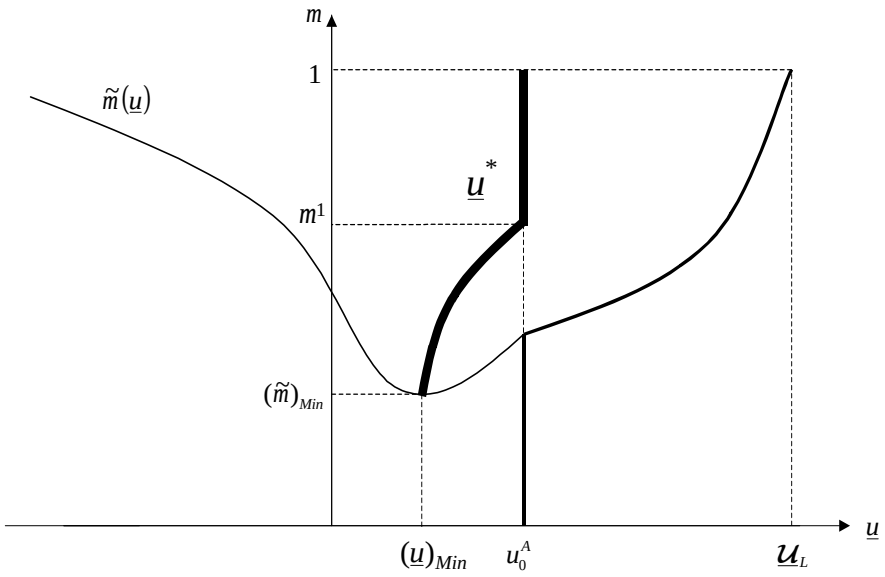


Figure 2: The optimal level $\underline{u}^{*}$

By simple observation, we can see another result which is linked to some observed legal punishments.

Proposition 5 *When the probability of facing a high risk agent is relatively high, the optimal level of minimal utility $\underline{u}^*$ is below the utility that corresponds to the nullity of the contract.*

We know that the locus $\underline{u}^*$ starts at the point where $\underline{u}^* = \widehat{u}^A_{L} = \underline{\mathcal{U}}_L$ and then decreases with μ, converging to the point where $\underline{u}^* = (\underline{u})_{Min}$. Therefore there must exist a value $\mu^1 > (\widetilde{\mu})_{Min}$ such that $\underline{u}^*(\mu^1) = u_0^A$. So, when the likelihood of facing a high risk policyholder is sufficiently high and above μ^1, this type of agent has an optimal $\underline{u}^*$ lower than u_0^A. Recall that this is the utility of a non-insured agent that has an accident. But u_0^A is also the level of utility that an insured gets if the company rescinds the contract. This is a widespread punishment for the kind of fraud that we analyzed in this paper. So when the probability of facing a high risk agent is above μ^1, from his personal point of view, the law should allow harder punishments than the nullity of his contract. As we mentioned in the Introduction, we do observe some variability in the minimal level of utilities $\underline{u}$ between different legislations. In some countries, the legal limit is set exclusively by the nullity of the contract while, in others, there is also a monetary fine for false declarations at the underwriting stage.

Finally we compare the optimal level of minimal utility $\underline{u}^*$ to the social optimum. In order to do that, we adopt an utilitarian approach. We define the social welfare $\mathcal{W}$ as the sum of the profit of the insurer and the expected utilities of both types of policyholders. It is straightforward to prove the last proposition of this paper.

Proposition 6 *The minimal level of utility $\underline{u}^*$ is not socially optimal. An utilitarian social authority should impose a lower legal limit $\underline{u}^{\mathcal{W}}$.*

Although the level of utility $\underline{u}^*$ represents a real punishment for a high risk agent in terms of loss of utility, this level is above the legal limit $\underline{u}^{\mathcal{W}}$ that should be imposed by an utilitarian social planner. The reason is that $\underline{u}^*$ does not completely internalize the insurer's expected profit as $\underline{u}^{\mathcal{W}}$ does. The social welfare would be higher with lower levels of minimal utility to set to defrauders, as the Beckerian approach points out.

5 Conclusion

We analyzed a Principal-Agent model of an insurer who faces an adverse selection problem. He is unable to observe if his client has a high risk or

a low risk of having an accident. At the underwriting of the contract, the insurer requests the client to declare his risk. After that, the former can costly audit the truthfulness of this announcement. If the audit confirms a false declaration, the insurer is legally allowed to punish the defrauder. We characterized the efficient contracts when this punishment is bounded from above by a legal restriction. These contracts are separating and imply, on the one hand, that a high risk agent is fully insured and may receive a positive informational rent. On the other hand, a low risk one is always partially insured at his reservation level of expected utility. These contracts include a probability of audit but only for a low risk report. Concerning potential fraud, they also specify that the punishment for a defrauder is set at the legal maximum.

Then, we did some comparative statics on the efficient contracts and on the agent's utility. The most important result of this paper concerns the legal limit to a defrauder's punishment. We prove that there exists a unique value of this legal limit that maximizes the expected utility of a high risk type. Facing this particular value of the legal limit to a defrauder's punishment, the insurer will effectively audit a low risk report. We also show that this particular value increases with the probability of facing a high risk policyholder. Therefore, when this probability is sufficiently high, the nullity of the contract is not enough. From the point of view of a potential defrauder, the law should allow harder sanctions. This is an striking result because the nullity of the contract is a common sanction for this kind of fraud in the USA and in some European countries.

Our framework seems too restrictive to apply for empirical issues and more policy oriented considerations because two criticisms can be made. The first concerns the Principal-Agent framework because it yields to a monopolistic representation of the insurance market. We agree that it is difficult to argue that, either in the USA or in Europe, the most important branches of the insurance industry are monopolies. Nevertheless, our model explains better than the perfect competition setting, some stylized facts specially the insurance industry's lobby for legal changes towards the criminalization of insurance fraud. In a competitive insurance market model, the companies are indifferent between any level of legal limit for punishments because no matter its value, they do not earn positive profits. So, as lobby is costly, in such model it is not worth for insurers to engage in this legislative activity. But, as this was certainly not the case at least in the USA, we can doubt about the validity of the competitive insurance market model for our

purposes. The second criticism focuses on another result that contradicts many practitioners observation, namely that our model does not give rise to fraud in equilibrium. In a more complicated framework where, for example, the insurer could not commit to the probability of audit, it has been shown elsewhere that it is not optimal for the latter to punish defrauders at the maximum legal level. Thus fraud occurs in equilibrium. Nevertheless, also in that case, the most important result of our paper would remain qualitatively unaltered. Defrauders will still have an optimal level of legal limit for punishments that will yield to non-negligeable sanctions for them, although lower than the one that we found here.

In spite of these criticisms, we believe that our model and its results can help to answer the following question: to what extent a change in the legislation towards more punishment for insurance defrauders will receive the approval of a majority? In that sense, we think that our approach can be generalized and pursued towards a political economy theory of anti-fraud legislation in insurance markets. In that perspective, our model could be extended to consider restrictions in other components of the insurance policy, as it was the case for the approval of the polemic Proposition 103 in California in 1988. Also we do not consider any externality problem here. In that case, low risk agents would take in account that they may have an accident with a defrauder which finally will not be covered. So, at the moment to vote for the limit to punishments, they will probably choose lower sanctions than the nullity of the contract. These may be some promising theoretical lines of research.

References

[1] ALFA (1999) "Fichiers Thématiques", n° 5.

[2] Baron, D. and Besanko, D. (1984) "Regulation, asymmetric information, and auditing", *RAND Journal of Economics*, **15**, 447-470.

[3] Becker, G. (1968) "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, **76**, 169-217.

[4] Bond, E. and Crocker, K. (1997) "Hardball and the soft touch: The economics of optimal insurance contracts with costly state verification and endogenous monitoring costs", *Journal of Public Economics*, **63**, 239-264.

[5] Border, K. C. and Sobel J. (1987) "Samourai Accountant: A Theory of Auditing and Plunder", *Review of Economic Studies*, **54**, 525-540.

[6] Coalition Against Insurance Fraud (1997) "Four Faces: Why some Americans do - and do not - tolerate insurance fraud" in *http://www.insurancefraud.org/.*

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

[8] Dixit, A. (2000) "Adverse Selection and Insurance with *Uberrima Fides*", Manuscript, Princeton University.

[9] Doherty, N. A. and Jung H. J. (1993) "Adverse Selection when Loss Severities Differ: First Best and Costly Equilibria", *Geneva Papers on Risk and Insurance Theory*, **18**, 173-182.

[10] Fagart, M.-C. and Picard, P. (1999) "Optimal Insurance Under Random Auditing", *The Geneva Papers on Risk and Insurance Theory*, **24**, 29-54.

[11] Florida Insurance Research Center (1991) "Automobile Insurance Fraud Study", University of Florida.

[12] Grossman, S. and Hart, O. (1983) "An analysis of the Principal-Agent problem", *Econometrica*, **51**, 7-45.

[13] Kansas Insurance Department (2000) "Kansas Auto Insurance...a necessity", in *http://www.ksinsurance.org/consumers/publications.*

[14] Mookherjee, D. and Png, I. (1989) "Optimal Auditing, Insurance and Redistribution", *The Quarterly Journal of Economics*, **104**, 399-415.

[15] Picard, P. (1996) "Auditing claims in the insurance market with fraud: The credibility issue", *Journal of Public Economics*, **63**, 27-56.

[16] Picard, P. (1999) "On the Design of Optimal Insurance Policies Under Manipulation of Audit Cost", forthcoming in *International Economic Review.*

[17] Reinganum, J. F. and Wilde, L. L. (1985) "Income Tax Compliance in a Principal-Agent Framework", *Journal of Public Economics*, **26**, 1-18.

[18] Rothschild, M. and Stiglitz, J. (1976) "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *Quarterly Journal of Economics*, **90**, 629-649.

[19] Stiglitz, J. (1977) "Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market", *Review of Economic Studies*, **44**, 407-430.

[20] Townsend, R. M. (1979) "Optimal Contracts and Competitive Markets with Costly State Verification", *Journal of Economic Theory*, **21**, 265-293

Appendix

Proof of Proposition 1

We rewrite the Stiglitz's model in the utility space, having defined the function $v \equiv u^{-1}(\,)$. The Lagrangian of the reduced problem of this model is as follows:[11]

$$\begin{aligned}\mathcal{L} = \quad & w - (\mu\pi_{_L} + (1-\mu)\pi_{_H})\ell \\ & -\mu\left[(1-\pi_{_L})v(u_{_L}^{NA}) + \pi_{_L}v(u_{_L}^{A})\right] - (1-\mu)v(u_{_H}) \\ & +\lambda_1\left[(1-\pi_{_L})u_{_L}^{NA} + \pi_{_L}u_{_L}^{A} - \underline{\mathcal{U}}_{_L}\right] \\ & +\alpha_1\left[u_{_H} - \underline{\mathcal{U}}_{_H}\right] \\ & +\lambda_2\left[u_{_H} - (1-\pi_{_H})u_{_L}^{NA} - \pi_{_H}u_{_L}^{A}\right]\end{aligned}$$

where λ_1 and λ_2 are the multipliers associated to the equality constraints and α_1, to the inequality one. The first-order conditions are

$$\left\{\begin{array}{lll} \frac{\partial\mathcal{L}}{\partial u_H} = & -(1-\mu)v'(\widehat{u}_{_H}) + \alpha_1 + \lambda_2 = 0 & FOC(1) \\ \frac{\partial\mathcal{L}}{\partial u_L^A} = & -\mu\pi_{_L}v'(\widehat{u}_{_L}^{A}) + \pi_{_L}\lambda_1 - \pi_{_H}\lambda_2 = 0 & FOC(2) \\ \frac{\partial\mathcal{L}}{\partial u_L^{NA}} = & -\mu(1-\pi_{_L})v'(\widehat{u}_{_L}^{NA}) + (1-\pi_{_L})\lambda_1 - (1-\pi_{_H})\lambda_2 = 0 & FOC(3) \\ & \alpha_1\left[\widehat{u}_{_H} - \underline{\mathcal{U}}_{_H}\right] = 0 \qquad \alpha_1 \geq 0 & CSC(1) \\ & (1-\pi_{_L})\widehat{u}_{_L}^{NA} + \pi_{_L}\widehat{u}_{_L}^{A} = \underline{\mathcal{U}}_{_L} & IR(L) \\ & (1-\pi_{_H})\widehat{u}_{_L}^{NA} + \pi_{_H}\widehat{u}_{_L}^{A} = u_{_H} & IC(H) \end{array}\right.$$

>From $FOC(2)$ and $FOC(3)$, we find

$$\lambda_2 = \frac{(1-\pi_{_L})\pi_{_L}\,\mu\left[v'(\widehat{u}_{_L}^{NA}) - v'(\widehat{u}_{_L}^{A})\right]}{\triangle\pi} \tag{1}$$

where $\triangle\pi \equiv \pi_{_H} - \pi_{_L}$. Also from $FOC(1)$

$$\begin{aligned}\alpha_1 \;&=\; (1-\mu)v'(\widehat{u}_{_H}) - \lambda_2 > 0 \qquad (2)\\ &=\; (1-\mu)v'(\widehat{u}_{_H}) - \frac{(1-\pi_{_L})\pi_{_L}\mu\left[v'(\widehat{u}_{_L}^{NA}) - v'(\widehat{u}_{_L}^{A})\right]}{\triangle\pi}\end{aligned}$$

[11] The reduced program considers only the binding constraints at the optimum.

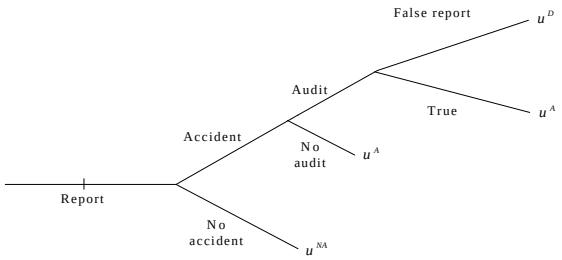
Next, we characterize the threshold of exclusion for a low risk individual. This threshold is given by the solution $\alpha_1 = 0$ and $\widehat{u}_{_H} = \underline{\mathcal{U}}_{_H}$. When this is the case, by construction $\widehat{u}_{_L}^{NA} = u_0^{NA}$ and $\widehat{u}_{_L}^{A} = u_0^{A}$. After some manipulations of (2), we obtain the threshold

$$0 < \widehat{\mu} = \frac{v'(\underline{\mathcal{U}}_{_H})}{v'(\underline{\mathcal{U}}_{_H}) + \frac{(1-\pi_{_L})\pi_{_L}\left[v'(u_0^{NA}) - v'(u_0^{A})\right]}{\Delta\pi}} < 1$$

If $\mu > \widehat{\mu}$ then $\alpha_1 = 0$ and $\widehat{u}_{_H} > \underline{\mathcal{U}}_{_H}$. Also $u_0^{NA} > \widehat{u}_{_L}^{NA} > \widehat{u}_{_L}^{A} > u_0^{A}$ and $\mathcal{U}_{_L} = \underline{\mathcal{U}}_{_L}$. Applying the Implicit Function Theorem and differentiating the first-order conditions with respect to μ, it is straightforward to verify that the utilities $\widehat{u}_{_L}^{A}$ and $\widehat{u}_{_H}$ are increasing functions of μ ■

Proof of Lemma 1 and Proposition 2

The timing of the game we are analyzing is the following:



In our setting, a contract $\mathcal{C}$ has the following shape

$$\mathcal{C} = \left\{ \ \widetilde{j} \in \{L, H\} \ \ \rightarrow \ \ \left(u_{\widetilde{j}}^{NA}, u_{\widetilde{j}}^{A}, \gamma_{\widetilde{j}}, u_{\widetilde{j}}^{D}\right) \ \right\}$$

where all final utilities must verify

$$\left\{ \begin{array}{l} u^{NA} \geq \underline{u} \\ u^{A} \geq \underline{u} \\ u^{D} \geq \underline{u} \end{array} \right. \tag{3}$$

and the expected utilities

$$\mathcal{U}_j(\mathcal{C}) \geq \underline{\mathcal{U}}_j \tag{4}$$

We also impose that an insured that is not audited must obtain the same utility than an insured that has been audited and the audit has confirmed his announcement.

Therefore, for any type $j \in \{L, H\}$, two potential reporting behavior $\widetilde{j}(j)$ may arise as a response to the contract $\mathcal{C}$ offered by the insurer. Either the type j reports truthfully $\widetilde{j}(j) = j$ because

$$(1-\pi_{_j})u_{_j}^{NA}+\pi_{_j}u_{_j}^{A} \geq (1-\pi_{_j})u_{_{j'}}^{NA}+\pi_{_j}\left(\gamma_{_{j'}}u_{_{j'}}^{D}+(1-\gamma_{_{j'}})u_{_{j'}}^{A}\right) \qquad j' \neq j \tag{5}$$

or he misreports $\widetilde{j}(j) \neq j$ because

$$(1-\pi_{_j})u_{_{j'}}^{NA}+\pi_{_j}\left(\gamma_{_{j'}}u_{_{j'}}^{D}+(1-\gamma_{_{j'}})u_{_{j'}}^{A}\right) > (1-\pi_{_j})u_{_j}^{NA}+\pi_{_j}u_{_j}^{A} \qquad j' \neq j \tag{6}$$

Proof of Lemma 1

We will show that a contract setting $\gamma_{_{\widetilde{j}}} > 0$ and $u_{_{\widetilde{j}}}^{D} > u_{_{\widetilde{j}}}^{A} \geq \underline{u}$ cannot be an efficient one. Assume first that the type j misreports. Therefore, the insurer's expected profit depends on the reporting behavior of type j'. If type j' misreports, this profit is

$$\begin{aligned}
\underset{j}{\mathbb{E}}\,\Pi = \; & p_{_j}\Big[\omega-\pi_{_j}\ell-(1-\pi_{_j})v(u_{_{j'}}^{NA}) \\
& -\pi_{_j}\left(\gamma_{_{j'}}v(u_{_{j'}}^{D})+(1-\gamma_{_{j'}})v(u_{_{j'}}^{A})+c(\gamma_{_{j'}})\right)\Big] \\
& +(1-p_{_j})\Big[\omega-\pi_{_{j'}}\ell-(1-\pi_{_{j'}})v(u_{_j}^{NA}) \\
& -\pi_{_{j'}}\left(\gamma_{_j}v(u_{_j}^{D})+(1-\gamma_{_j})v(u_{_j}^{A})+c(\gamma_{_j})\right)\Big]
\end{aligned} \tag{7}$$

where $p_{_j}$ is the probability of facing a type j. On the other hand, if type j' reports truthfully, the insurer's expected profit is

$$\begin{aligned}\underset{j}{\mathbb{E}}\,\Pi = \; & p_{_j}\Big[\omega-\pi_{_j}\ell-(1-\pi_{_j})v(u^{NA}_{j'}) \\ & -\pi_{_j}\Big(\gamma_{_{j'}}v(u^{D}_{j'})+(1-\gamma_{_{j'}})v(u^{A}_{j'})+c(\gamma_{_{j'}})\Big)\Big] \\ & +(1-p_{_j})\Big[\omega-\pi_{_{j'}}\ell-(1-\pi_{_{j'}})v(u^{NA}_{j'}) \\ & -\pi_{_{j'}}\Big(v(u^{A}_{j'})+c(\gamma_{_{j'}})\Big)\Big]\end{aligned} \tag{8}$$

For any strictly positive value of $\gamma_{_{j'}}$ and any strictly positive slack between $u^{D}_{j'}$ and $u^{A}_{j'}$, the insurer can always slightly decrease $\gamma_{_{j'}}$ by $d\gamma_{_{j'}}<0$ and $u^{D}_{j'}$ by $du^{D}_{j'}<0$ such that both reporting behavior remain unchanged and all constraints in (3) and (4) hold. Applying these modifications in (7) and (8),the insurer's expected profit increases. Therefore a contract like the initial one cannot be efficient. The same reasoning applies when, no matter the reporting behavior of type $j'\neq j$, the type j reports truthfully ■

Proof of Proposition 2

We will show that any couple of contracts that induce one type of agent or both types to misreport is strongly dominated, in the sense of the insurer's expected profit, by a truthful revealing one. Take an arbitrary couple of contracts

$$\mathcal{C}=\begin{cases}\widetilde{j}=L & \rightarrow \quad \{u^{NA}_{L},u^{A}_{L},\gamma_{_L},u^{D}_{L}\} \\ \widetilde{j}=H & \rightarrow \quad \{u^{NA}_{H},u^{A}_{H},\gamma_{_H},u^{D}_{H}\}\end{cases}$$

where $u^{A}_{\widetilde{j}}\geq u^{D}_{\widetilde{j}}$, all utilities verify (3) and, the expected utilities $\mathcal{U}_{_j}$, (4). Three different cases of reporting behavior $\widetilde{j}(j)$ may arise as a response to $\mathcal{C}$.

- Case 1: $\widetilde{j}(L)=H$ and $\widetilde{j}(H)=L$

  When both types misreport, their corresponding expected utilities verify

$$\begin{aligned}\mathcal{U}_{_{HL}}(\mathcal{C}) &= (1-\pi_{_L})u^{NA}_{H}+\pi_{_L}\left(\gamma_{_H}u^{D}_{H}+(1-\gamma_{_H})u^{A}_{H}\right) \\ &> (1-\pi_{_L})u^{NA}_{L}+\pi_{_L}u^{A}_{L}=\mathcal{U}_{_L}(\mathcal{C})\geq \underline{\mathcal{U}}_{_L}\end{aligned} \tag{9}$$

$$\begin{aligned}\mathcal{U}_{LH}(\mathcal{C}) &= (1-\pi_H)u_L^{NA} + \pi_H\left(\gamma_L u_L^D + (1-\gamma_L)u_L^A\right)\\ &> (1-\pi_H)u_H^{NA} + \pi_H u_H^A = \mathcal{U}_H(\mathcal{C}) \geq \underline{\mathcal{U}}_H\end{aligned} \tag{10}$$

and the insurer's expected profit is

$$\begin{aligned}\mathop{\mathbb{E}}_{j}\Pi(\mathcal{C}) = \;& \mu\left[\omega - \pi_L\ell - (1-\pi_L)v(u_H^{NA})\right.\\ & \left.-\pi_L\left(\gamma_H v(u_H^D) + (1-\gamma_H)v(u_H^A) + c(\gamma_H)\right)\right]\\ & +(1-\mu)\left[\omega - \pi_H\ell - (1-\pi_H)v(u_L^{NA})\right.\\ & \left.-\pi_H\left(\gamma_L v(u_L^D) + (1-\gamma_L)v(u_L^A) + c(\gamma_L)\right)\right]\end{aligned} \tag{11}$$

The insurer can offer a new couple of contracts

$$\mathcal{C}' = \begin{cases} \widetilde{j} = L & \rightarrow \quad \left\{u_H^{NA}, \gamma_H u_H^D + (1-\gamma_H)u_H^A, 0, \underline{u}\right\}\\ \widetilde{j} = H & \rightarrow \quad \left\{u_L^{NA}, \gamma_L u_L^D + (1-\gamma_L)u_L^A, 0, \underline{u}\right\}\end{cases}$$

Take $j = L$. If $\widetilde{j}(L) = L$

$$\begin{aligned}\mathcal{U}_L(\mathcal{C}') &= (1-\pi_L)u_H^{NA} + \pi_L\left(\gamma_H u_H^D + (1-\gamma_H)u_H^A\right)\\ &> (1-\pi_L)u_L^{NA} + \pi_L u_L^A && \text{from (9)}\\ &\geq (1-\pi_L)u_L^{NA} + \pi_L(\gamma_L u_L^D + (1-\gamma_L)u_L^A) && \text{from Lemma 1}\\ &= \mathcal{U}_{HL}(\mathcal{C}') && \text{the expected utility after a misreport } \widetilde{j}(L) = H\end{aligned}$$

As, facing the initial couple of contracts $\mathcal{C}$, the expected utility verified the individual rationality constraint, this is also the case here. So finally, $\widetilde{j}(L) = L$. Next take $j = H$. If $\widetilde{j}(H) = H$

$$\begin{aligned}\mathcal{U}_H(\mathcal{C}') &= (1-\pi_H)u_L^{NA} + \pi_H\left(\gamma_L u_L^D + (1-\gamma_L)u_L^A\right)\\ &> (1-\pi_H)u_H^{NA} + \pi_H u_H^A && \text{from (10)}\\ &\geq (1-\pi_H)u_H^{NA} + \pi_H(\gamma_H u_H^D + (1-\gamma_H)u_H^A) && \text{from Lemma 1}\\ &= \mathcal{U}_{LH}(\mathcal{C}') && \text{the expected utility after a misreport } \widetilde{j}(H) = L\end{aligned}$$

The same comment applies here concerning the type H individual ra-

tionality constraint. So $\widetilde{j}(H) = H$. This new couple of contracts $\mathcal{C}'$ induces truthful reporting of both types. The insurer's expected profit is now

$$\begin{aligned}
\mathop{\mathbb{E}}_{j}\Pi(\mathcal{C}') = \; & \mu\left[\omega - \pi_{_L}\ell - (1-\pi_{_L})v(u_{_H}^{NA})\right. \\
& \left.-\pi_{_L}v(\gamma_{_H}u_{_H}^{D} + (1-\gamma_{_H})u_{_H}^{A})\right] \\
& +(1-\mu)\left[\omega - \pi_{_H}\ell - (1-\pi_{_H})v(u_{_L}^{NA})\right. \\
& \left.-\pi_{_H}v(\gamma_{_L}u_{_L}^{D} + (1-\gamma_{_L})u_{_L}^{A})\right] \\
> \; & \mathop{\mathbb{E}}_{j}\Pi(\mathcal{C}) \qquad \text{because the function } v \text{ is strictly convex.}
\end{aligned}$$

Therefore the insurer strictly prefers to offer the new couple of contracts $\mathcal{C}'$.

- Case 2: $\widetilde{j}(L) = L$ and $\widetilde{j}(H) = L$

  When only a high risk agent misreports, the expected utilities now verify

$$\begin{aligned}
\mathcal{U}_{_L}(\mathcal{C}) &= (1-\pi_{_L})u_{_L}^{NA} + \pi_{_L}u_{_L}^{A} \\
&\geq (1-\pi_{_L})u_{_H}^{NA} + \pi_{_L}\left(\gamma_{_H}u_{_H}^{D} + (1-\gamma_{_H})u_{_H}^{A}\right) = \mathcal{U}_{_{HL}}(\mathcal{C})
\end{aligned} \tag{12}$$

$$\begin{aligned}
\mathcal{U}_{_{LH}}(\mathcal{C}) &= (1-\pi_{_H})u_{_L}^{NA} + \pi_{_H}\left(\gamma_{_L}u_{_L}^{D} + (1-\gamma_{_L})u_{_L}^{A}\right) \\
&> (1-\pi_{_H})u_{_H}^{NA} + \pi_{_H}u_{_H}^{A} = \mathcal{U}_{_H}(\mathcal{C}) \geq \underline{\mathcal{U}}_{_H}
\end{aligned} \tag{13}$$

and the insurer's expected profit is

$$\begin{aligned}
\mathop{\mathbb{E}}_{j}\Pi(\mathcal{C}) = \; & \mu\left[\omega - \pi_{_L}\ell - (1-\pi_{_L})v(u_{_L}^{NA}) - \pi_{_L}\left(v(u_{_L}^{A}) + c(\gamma_{_L})\right)\right] \\
& +(1-\mu)\left[\omega - \pi_{_H}\ell - (1-\pi_{_H})v(u_{_L}^{NA})\right. \\
& \left.-\pi_{_H}\left(\gamma_{_L}v(u_{_L}^{D}) + (1-\gamma_{_L})v(u_{_L}^{A}) + c(\gamma_{_L})\right)\right]
\end{aligned} \tag{14}$$

The insurer can offer the new couple of contracts

$$\mathcal{C}' = \begin{cases} \widetilde{j} = L & \rightarrow \quad \left\{u_{_L}^{NA}, u_{_L}^{A}, \gamma_{_L}, u_{_L}^{D} - \epsilon\right\} \\ \widetilde{j} = H & \rightarrow \quad \left\{u_{_L}^{NA}, \gamma_{_L}u_{_L}^{D} + (1-\gamma_{_L})u_{_L}^{A}, 0, \underline{u}\right\} \end{cases}$$

where $\epsilon > 0$. Take $j = L$. If $\widetilde{j}(L) = L$

$$\begin{aligned}
\mathcal{U}_{L}(\mathcal{C}') &= (1-\pi_{L})u_{L}^{NA} + \pi_{L}u_{L}^{A} \\
&\geq (1-\pi_{L})u_{L}^{NA} + \pi_{L}(\gamma_{L}u_{L}^{D} + (1-\gamma_{L})u_{L}^{A}) \quad \text{from Lemma 1} \\
&= \mathcal{U}_{HL}(\mathcal{C}') \quad \text{the expected utility after a misreport } \widetilde{j}(L) = H
\end{aligned}$$

So finally $\widetilde{j}(L) = L$ because $\mathcal{U}_{L}(\mathcal{C}') \geq \underline{\mathcal{U}}_{L}$. Next take $j = H$. If $\widetilde{j}(H) = H$

$$\begin{aligned}
\mathcal{U}_{H}(\mathcal{C}') &= (1-\pi_{H})u_{L}^{NA} + \pi_{H}\left(\gamma_{L}u_{L}^{D} + (1-\gamma_{L})u_{L}^{A}\right) \\
&> (1-\pi_{H})u_{L}^{NA} + \pi_{H}\left(\gamma_{L}(u_{L}^{D} - \epsilon) + (1-\gamma_{L})u_{L}^{A}\right) \\
&= \mathcal{U}_{LH}(\mathcal{C}') \quad \text{the expected utility after a misreport } \widetilde{j}(H) = L
\end{aligned}$$

so $\widetilde{j}(H) = H$ because $\mathcal{U}_{LH}(\mathcal{C}) \geq \underline{\mathcal{U}}_{H}$. This new couple of contracts $\mathcal{C}'$ induces truthful reporting of both types. The insurer's expected profit is now

$$\begin{aligned}
\mathop{\mathbb{E}}_{j}\Pi(\mathcal{C}') = \quad & \mu\left[\omega - \pi_{L}\ell - (1-\pi_{L})v(u_{L}^{NA}) - \pi_{L}\left(v(u_{L}^{A}) + c(\gamma_{L})\right)\right] \\
& +(1-\mu)\left[\omega - \pi_{H}\ell - (1-\pi_{H})v(u_{L}^{NA})\right. \\
& \left.-\pi_{H}(v(\left(\gamma_{L}u_{L}^{D} + (1-\gamma_{L})u_{L}^{A}\right)) + c(\gamma_{L}))\right] \\
> \quad & \mathop{\mathbb{E}}_{j}\Pi(\mathcal{C}) \qquad \text{because the function } v \text{ is strictly convex.}
\end{aligned}$$

Again the insurer strictly prefers to offer the new couple of contracts $\mathcal{C}'$.

- Case 3: $\widetilde{j}(L) = H$ and $\widetilde{j}(H) = H$

  When only a low risk agent misreports, the proof is similar to the previous one.

Therefore, for any arbitrary couple of contracts that induces a misreport, the insurer strictly prefers to offer a truthful revealing couple of contracts. ■

Proof of Lemma 2

An incentive compatible contract must verify

$$\mathcal{U}_j \geq \mathcal{U}_{j'j} = (1-\pi_j)u_{j'}^{NA} + \pi_j[\gamma_{j'}u_{j'}^{D} + (1-\gamma_{j'})u_{j'}^{A}]$$

Assume an efficient contract where $u_{j'}^{D} > \underline{u}$. The insurer can lower $u_{j'}^{D}$ and $\gamma_{j'}$ so as to maintain $\mathcal{U}_{j'j}$ constant. By doing so, the insurer's expected profit increases because the expected audit cost decreases. So the initial contract could not be efficient ■

Proof of Lemma 3

Assume that an efficient contract sets $u_{H}^{NA} > u_{H}^{A}$. The insurer can decrease u_{H}^{NA} and increase u_{H}^{A} such that the policyholder's expected utility $\mathcal{U}_{H}$ is unaffected. By construction, the new couple of contracts verifies all constraints and the insurer's expected profit increases. Therefore the initial contract could not be efficient ■

Proof of Lemma 4

To prove this lemma, we need the following intermediary result.

Lemma 7 *If $IC(L)$ binds then $IC(H)$ is slack.*

Proof. Assume that $IC(L)$ binds so $\mathcal{U}_{L} = (1-\pi_{L}\gamma_{H})u_{H} + \pi_{L}\gamma_{H}\underline{u}$. Then, either $\gamma_{H} = 0$ or $\gamma_{H} > 0$.

1) If $\gamma_{H} = 0$, $\mathcal{U}_{L} = u_{H}$ and either $\gamma_{L} = 0$ or $\gamma_{L} > 0$. If $\gamma_{L} = 0$, this is the Stiglitz framework where $IC(L)$ does not bind in equilibrium. If $\gamma_{L} > 0$ then

$$\gamma_{L}\underline{u} + (1-\gamma_{L})u_{L}^{A} \leq u_{L}^{A}$$

by Lemma 1. So

$$\mathcal{U}_{LH} < (1-\pi_{H})u_{L}^{NA} + \pi_{H}u_{L}^{A}$$

But, by $NOI(L)$ and $\pi_{H} > \pi_{L}$

$$(1-\pi_{H})u_{L}^{NA} + \pi_{H}u_{L}^{A} < (1-\pi_{L})u_{L}^{NA} + \pi_{L}u_{L}^{A} = \mathcal{U}_{L}$$

Then $u_{H} = \mathcal{U}_{L} > \mathcal{U}_{LH}$ and therefore $IC(H)$ is slack.

2) If $\gamma_{H} > 0$, by Lemma 1 we have that

$$u_{H} > (1 - \pi_{L}\gamma_{H})u_{H} + \pi_{L}\gamma_{H}\underline{u} = \mathcal{U}_{L}$$

But as

$$\gamma_{L}\underline{u} + (1 - \gamma_{L})u_{L}^{A} < u_{L}^{A}$$

and $\pi_{H} > \pi_{L}$ then $\mathcal{U}_{L} > \mathcal{U}_{LH}$ so $IC(H)$ is slack ■

So assume that $IC(L)$ binds at the optimum. Then

$$\mathcal{U}_{L} = (1 - \pi_{L}\gamma_{H})u_{H} + \pi_{L}\gamma_{H}\underline{u}$$

>From the previous lemma, $\forall\gamma_{H}\ u_{H} > \underline{\mathcal{U}}_{H}$. Hence, we can decrease u_{H} such that $IC(H)$ and $IR(H)$ remain and $IC(L)$ becomes slack. As the contract for the low risk is unchanged, $IR(L)$ remains. The incentive properties of the initial contract remain but the insurer's expected profit increases. Hence the initial contract could not be efficient. So $IC(L)$ must be slack at the optimum■

Next, assume that an efficient contract sets $\gamma_{H} > 0$. As $IC(L)$ is slack, then

$$\mathcal{U}_{L} > (1 - \pi_{L}\gamma_{H})u_{H} + \pi_{L}\gamma_{H}\underline{u}$$

Then, the insurer can slightly decrease γ_{H} such that $IC(L)$ remains slack. By construction, the new contract verifies all constraints but the insurer's expected profit increases. Then $\gamma_{H} > 0$ cannot be optimal■

Finally, let an efficient contract yielding to $IC(H)$ slack. Then, either $\gamma_{L} > 0$ or $\gamma_{L} = 0$.

1. if $\gamma_{L} = 0$ it is the Stiglitz framework. So $IC(H)$ binds with equality, which is a contradiction.

2. so $\gamma_{L} > 0$. Then the insurer can decrease γ_{L} and by doing so increase $\mathcal{U}_{LH}$ but such that $IC(H)$ remains slack. By construction, the new contract verifies all constraints and the insurer's expected profit increases. Then an efficient contract could not verify $IC(H)$ slack ■

Proof of Lemma 5

Assume that $IR(L)$ is slack at the optimum. The insurer can slightly decrease u_{L}^{NA} and u_{L}^{A}. In that case, du_{L}^{NA} and du_{L}^{A} can be found such that $IC(L)$ and

$IR(L)$ are still slack and $IC(H)$ holds. But therefore it is straightforward to verify that the insurer's expected profit increases. Hence an efficient contract can not set $IR(L)$ slack ■

Proof of Lemma 6

Assume that $NOI(L)$ binds and $\gamma_{_L} < \tau \equiv \frac{\triangle\pi}{\pi_{_H}(1-\pi_{_L})}$ in equilibrium. The insurer can slightly increase $u_{_L}^{NA}$ and decrease $u_{_L}^{A}$ such that $IR(L)$ still holds. Applying this changes, the right hand side of $IC(H)$ becomes

$$(1-\pi_{_H})du_{_L}^{NA} + \pi_{_H}(1-\gamma_{_L})du_{_L}^{A}$$

Assume that $IR(H)$ binds. In order to maintain the equality in $IC(H)$, the insurer can also change $\gamma_{_L}$ such that

$$[\pi_{_H}(1-\gamma_{_L}) - \frac{\pi_{_L}(1-\pi_{_H})}{(1-\pi_{_L})}]du_{_L}^{A} = \pi_{_H}(u_{_L}^{A} - \underline{u})d\gamma_{_L}$$

As $\gamma_{_L} < \tau$ and $du_{_L}^{A} < 0, d\gamma_{_L} < 0$. By doing that changes the insurer's expected profit increases. Hence an efficient contract can not be like this. The intuition follows identically if we assume $IR(H)$ slack. The same proof also holds if we assume an efficient contract setting $NOI(L)$ slack and $\gamma_{_L} \geq \tau$ ■

The efficient contracts

In order to completely characterize the contracts offered by the insurer, we proceed in two steps. First of all, we obtain all possible types of efficient contracts. Next we show, in the $(\underline{u}, \mu)$ space, the parametric regions where each type of contract dominates. Recall that we have restricted $\underline{u}$ to verify $\underline{u} \leq \widehat{u}_{_L}^{A}$.

We have to find the solution of the following problem

$$\begin{aligned}
\underset{u_L^{NA},u_L^A,\gamma_L,u_H}{Max} \quad & \mu\left[\omega-\pi_L\ell-(1-\pi_L)v(u_L^{NA})-\pi_L\left(v(u_L^A)+c(\gamma_L)\right)\right] \\
& +(1-\mu)[\omega-\pi_H\ell-v\left(u_H\right)] \\
& \textit{subject to} \\
& (1-\pi_L)u_L^{NA}+\pi_L u_L^A=\underline{\mathcal{U}}_L && IR_L \\
& u_H\geq\underline{\mathcal{U}}_H && IR_H \\
& u_H=(1-\pi_H)u_L^{NA}+\pi_H((1-\gamma_L)u_L^A+\gamma_L\underline{u}) && IC_H \\
& \gamma_L\geq 0
\end{aligned}$$

whose Lagrangian is:

$$\begin{aligned}
\mathcal{L}= \quad & w-(\mu\pi_L+(1-\mu)\pi_H)\ell \\
& -\mu((1-\pi_L)v(u_L^{NA})+\pi_L v(u_L^A))-(1-\mu)v(u_H)-\mu\pi_L c(\gamma_L) \\
& +\lambda_1[(1-\pi_L)u_L^{NA}+\pi_L u_L^A-\underline{\mathcal{U}}_L] \\
& +\lambda_2[u_H-(1-\pi_H)u_L^{NA}+\pi_H((1-\gamma_L)u_L^A+\gamma_L\underline{u})] \\
& +\alpha_1[u_H-\underline{\mathcal{U}}_H]+\alpha_2\gamma_L
\end{aligned}$$

where λ_1 and λ_2 are the (strictly positive) multipliers associated to the equality constraints and α_1 and α_2, to the inequality ones.[12] The system of first

[12] We know that λ_1 and λ_2 are strictly positive because they are in fact Kuhn-Tucker multipliers associated to inequalities that bind at the optimum.

order conditions is the following :

$$\mathcal{S}_1 = \begin{cases} \frac{\partial \mathcal{L}}{\partial u_L^{NA}} & = -\mu(1-\pi_L)v'(u_L^{NA}) + (1-\pi_L)\lambda_1 - (1-\pi_H)\lambda_2 = 0 & FOC(1) \\ \frac{\partial \mathcal{L}}{\partial u_L^{A}} & = -\mu\pi_L v'(u_L^{A}) + \pi_L\lambda_1 - \pi_H(1-\gamma_L)\lambda_2 = 0 & FOC(2) \\ \frac{\partial \mathcal{L}}{\partial u_H} & = -(1-\mu)v'(u_H) + \alpha_1 + \lambda_2 = 0 & FOC(3) \\ \frac{\partial \mathcal{L}}{\partial \gamma_L} & = -\mu\pi_L c'(\gamma_L) + \pi_H\lambda_2\left(u_L^A - \underline{u}\right) + \alpha_2 = 0 & FOC(4) \\ & \underline{\mathcal{U}}_L = (1-\pi_L)u_L^{NA} + \pi_L u_L^A & IR(L) \\ & u_H = (1-\pi_H)u_L^{NA} + \pi_H((1-\gamma_L)u_L^A + \gamma_L\underline{u}) & IC(H) \\ & \alpha_1\left[u_H - \underline{\mathcal{U}}_H\right] = 0 \qquad \alpha_1 \geq 0 & CSC(1) \\ & \alpha_2\gamma_L = 0 \qquad \alpha_2 \geq 0 & CSC(2) \end{cases}$$

We may have, as a solution of $\mathcal{S}_1$, the Stiglitz case ($\gamma_L = 0$), characterized by $\alpha_2 > 0$ and the corresponding system of first-order conditions. Then, when the probability of audit is strictly positive, two solutions may arise.

Case A: $u_L^{NA} > u_L^A$ **and** $u_H > \underline{\mathcal{U}}_H$

>From the complementary-slackness conditions $CSC(1)$ and $CSC(2)$, $\alpha_1 = 0$ and $\alpha_2 = 0$. Therefore the system of first-order conditions becomes

$$\mathcal{S}_1^A = \begin{cases} (1-\mu)v'(u_H)[\triangle\pi - \pi_H\gamma_L(1-\pi_L)] = \mu\pi_L(1-\pi_L)[v'(u_L^{NA}) - v'(u_L^A)] \\ \pi_H(1-\mu)v'(u_H)\left(u_L^A - \underline{u}\right) = \mu\pi_L c'(\gamma_L) \\ u_H = (1-\pi_H)u_L^{NA} + \pi_H[(1-\gamma_L)u_L^A + \gamma_L\underline{u}] \\ \underline{\mathcal{U}}_L = (1-\pi_L)u_L^{NA} + \pi_L u_L^A \end{cases}$$

Case B: $u_{L}^{NA} > u_{L}^{A}$ **and** $u_{H} = \underline{\mathcal{U}}_{H}$

>From $CSC(2)$, $\alpha_2 = 0$. So the system of first-order conditions is in that case

$$\mathcal{S}_1^B = \begin{cases} [(1-\mu)v'(\underline{\mathcal{U}}_{H}) - \alpha_1][\triangle\pi - \pi_{H}\gamma_{L}(1-\pi_{L})] \\ \qquad = \mu\pi_{L}(1-\pi_{L})[v'(u_{L}^{NA}) - v'(u_{L}^{A})] \\ \pi_{H}[(1-\mu)v'(\underline{\mathcal{U}}_{H}) - \alpha_1]\left(u_{L}^{A} - \underline{u}\right) = \mu\pi_{L}c'(\gamma_{L}) \\ \underline{\mathcal{U}}_{H} = (1-\pi_{H})u_{L}^{NA} + \pi_{H}[(1-\gamma_{L})u_{L}^{A} + \gamma_{L}\underline{u}] \\ \underline{\mathcal{U}}_{L} = (1-\pi_{L})u_{L}^{NA} + \pi_{L}u_{L}^{A} \end{cases}$$

The frontiers of the parametric regions

We have to characterize the regions, in the $(\underline{u}, \mu)$ space, where each of the possible solutions are in fact optimal to implement. First we draw the locus of pairs $(\underline{u}, \mu)$ that characterize the Stiglitz case. This particular solution to our problem is in fact described by the kinky curve $\underline{u} = \widehat{u}_{L}^{A}(\mu)$.

The frontier between the cases A and B can be derived from one particular solution to $\mathcal{S}_1^B$, namely when $\alpha_1 = 0$. This solution gives, in addition to the control variables of the problem, a locus $\widetilde{\mu}(\underline{u})$ where the pairs of parameters $(\underline{u}, \mu)$ must lye on. When $\gamma_{L} > 0$, it is necessary that $\underline{u} \leq u_0^A$. Given that, we obtain the following results when $(\underline{u}, \mu) \in \widetilde{\mu}(\underline{u})$.

$$\lim_{\underline{u}\to u_0^A} \gamma_{L} = 0 \quad \text{and} \quad \lim_{\underline{u}\to u_0^A} \mu = \widehat{\mu}$$

Moreover, we find that

$$\lim_{\underline{u}\to-\infty} \gamma_{L} = 0\ , \quad \lim_{\underline{u}\to-\infty} u_{L}^{NA} = \lim_{\underline{u}\to-\infty} u_{L}^{A} = \underline{\mathcal{U}}_{L} \text{ and } \lim_{\underline{u}\to-\infty} \mu = 1$$

Next we compute

$$\lim_{\underline{u}\to u_0^A} \frac{d\widetilde{\mu}(\underline{u})}{d\underline{u}} = \frac{\frac{1}{\mu c''(0)}[\pi_{H}(1-\pi_{L})v'(\underline{\mathcal{U}}_{H})]^2}{\frac{\triangle\pi}{1-\pi_{L}}v'(\underline{\mathcal{U}}_{H}) + \pi_{L}\left[v'(u_0^{NA}) - v'(u_0^{A})\right]} > 0$$

The last step to visualize the shape of the locus $\widetilde{\mu}(\underline{u})$ is to see that

$$\frac{d\widetilde{\mu}(\underline{u})}{d\underline{u}} = 0 \Leftrightarrow \triangle\pi(1-\mu)v'(\underline{\mathcal{U}}_{H}) = 2\mu\pi_{L}[(1-\pi_{L})v''(u_{L}^{A}) + \pi_{L}v''(u_{L}^{NA})](u_{L}^{A} - \underline{u})$$

We denote by $(\underline{u})_{Min}$ the value of $\underline{u}$ that verifies this equality. The value of μ that solves the equation

$$\mu = \widetilde{\mu}\left((\underline{u})_{Min}\right)$$

is denoted by $(\widetilde{\mu})_{Min}$. Both values verify

$$(\underline{u})_{Min} < u_0^A$$

and

$$(\widetilde{\mu})_{Min} < \widehat{\mu}$$

■

Proof of Proposition 3

Existence of a local maximum of a high risk agent's expected utility

We have to find if there exists an optimal level of $\underline{u}$ for a high risk policyholder. In order to do that, we compute the partial derivative of $u_{_H}$ with respect to $\underline{u}$ when the Case A holds. To do that, we differentiate the system of first-order conditions $\mathcal{S}_1^A$ with respect to $\underline{u}$ and we apply the Implicit Function Theorem. We obtain the system

$$\mathcal{S}_2 = \begin{cases} \mu[\pi_{_L} v''(u_{_L}^{NA}) + (1-\pi_{_L}) v''(u_{_L}^A)]\frac{\partial u_{_L}^A}{\partial \underline{u}} - (1-\pi_{_L})\frac{\pi_{_H}}{\pi_{_L}}(1-\mu)v'(u_{_H})\frac{\partial \gamma_{_L}}{\partial \underline{u}} \\ +(1-\mu)v''(u_{_H})\frac{\partial u_{_H}}{\partial \underline{u}}\left(\frac{\triangle\pi - \pi_{_H}\gamma_{_L}(1-\pi_{_L})}{\pi_{_L}}\right) = 0 \\ \pi_{_H}(1-\mu)v'(u_{_H})\frac{\partial u_{_L}^A}{\partial \underline{u}} - \mu\pi_{_L} c\frac{\partial \gamma_{_L}}{\partial \underline{u}} \\ +\pi_{_H}(1-\mu)\left(u_{_L}^A - \underline{u}\right)v''(u_{_H})\frac{\partial u_{_H}}{\partial \underline{u}} = \pi_{_H}(1-\mu)v'(u_{_H}) \\ (1-\pi_{_H})\frac{\partial u_{_L}^{NA}}{\partial \underline{u}} + \pi_{_H}(1-\gamma_{_L})\frac{\partial u_{_L}^A}{\partial \underline{u}} - \frac{\partial u_{_H}}{\partial \underline{u}} - \pi_{_H}(u_{_L}^A - \underline{u})\frac{\partial \gamma_{_L}}{\partial \underline{u}} = -\pi_{_H}\gamma_{_L} \\ (1-\pi_{_L})\frac{\partial u_{_L}^{NA}}{\partial \underline{u}} + \pi_{_L}\frac{\partial u_{_L}^A}{\partial \underline{u}} = 0 \end{cases}$$

After some algebra, we find that

$$\frac{\partial u_{_H}}{\partial \underline{u}} = \frac{N}{D}$$

where the numerator N is equal to

$$
\begin{aligned}
&(1-\pi_{_L})\triangle\pi[\pi_{_H}(1-\mu)v'(u_{_H})]^2\\
&-\pi_{_L}(1-\pi_{_L})\pi_{_H}^2\mu[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})]\ (u_{_L}^A-\underline{u})(1-\mu)v'(u_{_H})\\
&-\pi_{_L}^2(1-\pi_{_L})\pi_{_H}\mu^2[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})]\ \gamma_{_L}c
\end{aligned}
$$

and the denominator D is equal to

$$
\begin{aligned}
&[\pi_{_H}(1-\pi_{_L})(1-\mu)v'(u_{_H})]^2\\
&-\mu^2\pi_{_L}^2c(1-\pi_{_L})[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})]\\
&-\mu(1-\mu)\pi_{_L}\left[\triangle\pi-\pi_{_H}\gamma_{_L}(1-\pi_{_L})\right]^2cv''(u_{_H})\\
&+2\ (1-\mu)^2(1-\pi_{_L})\pi_{_H}^2\left[\triangle\pi-\pi_{_H}\gamma_{_L}(1-\pi_{_L})\right]v'(u_{_H})v''(u_{_H})\ (u_{_L}^A-\underline{u})\\
&-\mu(1-\mu)(1-\pi_{_L})\pi_{_L}\pi_{_H}^2[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})]v''(u_{_H})(u_{_L}^A-\underline{u})^2
\end{aligned}
$$

We can have some information about the sign of this derivative. First of all, we analyze the sign of D. Dividing its expression by $v''(u_{_H})=\beta\left[v'(u_{_H})\right]^2>0$, we obtain

$$
\begin{aligned}
\tfrac{D}{v''(u_{_H})}=\ &\tfrac{[\pi_{_H}(1-\pi_{_L})(1-\mu)]^2}{\beta}\\
&-\mu^2\pi_{_L}^2(1-\pi_{_L})c\tfrac{(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})}{v''(u_{_H})}\\
&-\mu(1-\mu)\pi_{_L}c\left[\triangle\pi-\pi_{_H}\gamma_{_L}(1-\pi_{_L})\right]^2\\
&+2\ (1-\mu)^2(1-\pi_{_L})\pi_{_H}^2\left[\triangle\pi-\pi_{_H}\gamma_{_L}(1-\pi_{_L})\right]v'(u_{_H})(u_{_L}^A-\underline{u})\\
&-(1-\mu)(1-\pi_{_L})\pi_{_L}\mu\pi_{_H}^2[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})](u_{_L}^A-\underline{u})^2
\end{aligned}
$$

As $u_{_L}^{NA}>u_{_L}^A>u_0^A$ and $u_{_H}<\underline{\mathcal{U}}_{_L}$, it is straightforward to see that

$$
\tfrac{D}{v''(u_{_H})}<\ \tfrac{1}{\beta}+2v'(\underline{\mathcal{U}}_{_L})-c\tfrac{v''(u_0^A)}{v''(\underline{\mathcal{U}}_{_L})}
$$

So if

$$
c\geq\frac{v''(u_0^A)}{v''(\underline{\mathcal{U}}_{_L})}\left[\frac{1}{\beta}+2v'(\underline{\mathcal{U}}_{_L})\right]\equiv\widehat{c}_1
$$

then the denominator D is always strictly negative.

On one side, we know that the Case B is the left limit of the Case A when $\mu \geq \widehat{\mu}$. So when $(\underline{u}, \mu) \to \widetilde{\mu}(\underline{u})$, $\frac{\partial u_H}{\partial \underline{u}} > 0$. On the other side, when $\underline{u}$ increases so that the frontier between the Case A and the Stiglitz solution is attained, we have

$$\lim_{\underline{u} \to \widehat{u}_L^A} \frac{\partial u_H}{\partial \underline{u}} = \frac{1}{\frac{(1-\pi_L)}{\triangle \pi} - \mu \pi_L c \frac{\mu(1-\pi_L)\pi_L[(1-\pi_L)v''(\widehat{u}_L^A) + \pi_L v''(\widehat{u}_L^{NA})] + \triangle \pi^2 (1-\mu) v''(\widehat{u}_H)}{\triangle \pi (1-\pi_L)[\pi_H (1-\mu) v'(\widehat{u}_H)]^2}}$$

Although the sign of this expression is not straightforward, we can show another sufficient condition so that it is always negative. Let's restrict our attention to the following expression

$$\frac{(1-\pi_L)}{\triangle \pi} - \pi_L c \frac{\mu}{(1-\mu)} \frac{\triangle \pi v''(\widehat{u}_H)}{(1-\pi_L)[\pi_H v'(\widehat{u}_H)]^2} \tag{15}$$

As $v(u) = \frac{-\ln(1-u)}{\beta}$, $\frac{v''(\widehat{u}_H)}{v'(\widehat{u}_H)^2}$ is increasing in μ because $\frac{d\widehat{u}_H}{d\mu} > 0$. Moreover $\frac{\mu}{(1-\mu)}$ also increases with μ. As $\mu \geq \widehat{\mu}$, it is sufficient to set

$$c > \left(\frac{(1-\pi_L)}{\triangle \pi} \right)^3 \frac{v'(\underline{\mathcal{U}}_H)}{v''(\underline{\mathcal{U}}_H)} \pi_H^2 \left[v'(u_0^{NA}) - v'(u_0^A) \right] \equiv \widehat{c}_2$$

Denote $\widehat{c} = Max\{\widehat{c}_1, \widehat{c}_2\}$ and assume that $c > \widehat{c}$. This ensures that (15) is negative so $\lim_{\underline{u} \to \widehat{u}_L^A} \frac{\partial u_H}{\partial \underline{u}} < 0$. Therefore, as D is also always negative, for any $\mu > \widehat{\mu}$, there exists at least a value of $\underline{u}$ such that $N = 0$ and $\frac{\partial u_H}{\partial \underline{u}} = 0$. So u_H has at least one local maximum denoted by u_H^*. This particular value of $\underline{u}$ depends on all the parameters of the model ■

Uniqueness of the local maximum

We have to prove some intermediary results

Lemma 8 *When $N = 0$, γ_L and u_L^A are locally decreasing in $\underline{u}$.*

Proof. >From $\mathcal{S}_1^A$ we have

$$(1-\mu)v'(u_H)[\triangle \pi - \pi_H \gamma_L (1-\pi_L)] = \mu \pi_L (1-\pi_L)[v'(u_L^{NA}) - v'(u_L^A)] \tag{16}$$

$$\mu \pi_L c'(\gamma_L) = \pi_H (1-\mu) v'(u_H)(u_L^A - \underline{u}) \tag{17}$$

Denote $(\gamma_{_L})^*, (u_{_L}^{NA})^*, (u_{_L}^{A})^*$ and $(u_{_H})^*$ the values taken by the utilities when $N = 0$. Lets analyze the impact of an infinitesimal change $\epsilon > 0$ of $\underline{u}$. From the fact that $N = 0$, the resulting changes in $u_{_H}$ can be neglected at a first-order.

First we maintain fixed $(u_{_L}^{A})^*$. From (17)

$$\gamma_{_L} \simeq (\gamma_{_L})^* - \epsilon(c')^{-1}\left(\frac{\pi_{_H}(1-\mu)v'(u_{_H})}{\mu\pi_{_L}}\right)$$

As the function $c(\gamma_{_L})$ is positive and monotonic, the function $(c')^{-1}$ is always positive. Therefore $\gamma_{_L} < (\gamma_{_L})^*$. Now plugging this result in (16), the left-hand side increases. In order to re-establish the equality, the right-hand side must increase. As $\frac{\partial u_{_L}^{NA}}{\partial \underline{u}} = \frac{-\pi_{_L}}{1-\pi_{_L}}\frac{\partial u_{_L}^{A}}{\partial \underline{u}}$, the first-order change in the right-hand side is

$$-\mu\pi_{_L}[(1-\pi_{_L})v''(u_{_L}^{A}) + \pi_{_L}v''(u_{_L}^{NA})]\frac{\partial u_{_L}^{A}}{\partial \underline{u}}$$

Therefore $(u_{_L}^{A})^*$ must decrease. This change re-enforces the negative change in $(\gamma_{_L})^*$ ■

Lemma 9 *When $N = 0$, the audit probability $\gamma_{_L}$ is lower than $\frac{\tau}{2}$.*

Proof. >From $\mathcal{S}_2$, evaluated at $N = 0$, it is straightforward to show that

$$\begin{cases} \left.\frac{\partial \gamma_{_L}}{\partial \underline{u}}\right|_{N=0} = \frac{\pi_{_H}(1-\mu)v'(u_{_H})\triangle\pi}{\mu\pi_{_L}c''(\gamma_{_L})[2\pi_{_H}\gamma_{_L}(1-\pi_{_L})-\triangle\pi]} \\ \left.\frac{\partial u_{_L}^{A}}{\partial \underline{u}}\right|_{N=0} = \frac{2\pi_{_H}\gamma_{_L}(1-\pi_{_L})}{2\pi_{_H}\gamma_{_L}(1-\pi_{_L})-\triangle\pi} \end{cases}$$

By Lemma 8, both expression are negative. So $\gamma_{_L} < \frac{\triangle\pi}{2\pi_{_H}(1-\pi_{_L})} = \frac{\tau}{2}$ ■

Lemma 10 *If $\beta\ell < \frac{2}{3}\ln\frac{1-\pi_{_L}}{\pi_{_L}}$ then $(1-\pi_{_L})^2v'''(u_{_L}^{A}) > (\pi_{_L})^2v'''(u_{_L}^{NA})$ when $N = 0$.*

Proof. The expression $(1-\pi_{_L})^2v'''(u_{_L}^{A})-(\pi_{_L})^2v'''(u_{_L}^{NA})$ attains its minimum when $u_{_L}^{A} = u_0^{A}$ and $u_{_L}^{NA} = u_0^{NA}$. Therefore, if we assume that

$$(1-\pi_{_L})^2v'''(u_0^{A}) - (\pi_{_L})^2v'''(u_0^{NA}) > 0$$

this sign will remain. With our functional specifications, this assumption yields to

$$2\beta\left[\frac{(1-\pi_{_L})^2}{(1-u_0^A)^3}-\frac{\pi_{_L}{}^2}{(1-u_0^{NA})^3}\right]>0$$

where $u_0^A=1-e^{-\beta\omega}$ and $u_0^{NA}=1-e^{-\beta(\omega-\ell)\cdot}$. The condition

$$\beta\ell<\frac{2}{3}\ln\frac{1-\pi_{_L}}{\pi_{_L}}$$

is therefore immediate. When ℓ and $\pi_{_L}$ are sufficiently small, this condition holds ■

The last step to prove the uniqueness of the maximum level of $u_{_H}$ is to find the sign of $\frac{\partial^2 u_{_H}}{\partial \underline{u}^2}\Big|_{N=0}$. As $\frac{\partial u_{_H}}{\partial \underline{u}}=\frac{N}{D}$, $\frac{\partial^2 u_{_H}}{\partial \underline{u}^2}\Big|_{N=0}=\frac{1}{D}\frac{\partial N}{\partial \underline{u}}\Big|_{N=0}$. Moreover, as $c'(\gamma_{_L})=c\gamma_{_L}$

$$N=0\Leftrightarrow\triangle\pi(1-\mu)v'(u_{_H})=2\mu\pi_{_L}[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})](u_{_L}^A-\underline{u})$$

so

$$\begin{aligned}\frac{\partial N}{\partial \underline{u}}\Big|_{N=0}= \; & -2\mu\pi_{_L}\Big\{[(1-\pi_{_L})^2v'''(u_{_L}^A)-(\pi_{_L})^2v'''(u_{_L}^{NA})]\frac{\partial u_{_L}^A}{\partial \underline{u}}(u_{_L}^A-\underline{u})\\ & +[(1-\pi_{_L})v''(u_{_L}^A)+\pi_{_L}v''(u_{_L}^{NA})]\left(\frac{\partial u_{_L}^A}{\partial \underline{u}}-1\right)\Big\}\end{aligned}$$

Applying the lemmas 8 and 10, $\frac{\partial N}{\partial \underline{u}}\Big|_{N=0}>0$. As D is always negative, $\frac{\partial^2 u_{_H}}{\partial \underline{u}^2}<0$ when $N=0$. As at that point $u_{_H}$ is strictly concave, the critical point is a local maximum. So if there exists another critical point where $N=0$, it could not be a minimum. But as this is a necessary condition for the existence of another local maximum, the value $\underline{u}^*$ such that $\frac{\partial u_{_H}}{\partial \underline{u}}=0$ is unique and the maximum $u_{_H}^*$ is global. We conclude that, for each value of μ, there exist an unique level $\underline{u}=\underline{u}^*$ such that the utility of a high risk agent is maximal■

Some comparative statics on $\underline{u}^*$

We denote by $\underline{u}^*(\mu)$ the locus of values $\underline{u}$ that maximize $u_{_H}$. We do some comparative statics on $\underline{u}^*(\mu)$. From $\mathcal{S}_1^A$ we can derive the following system

denoted by $\mathcal{S}_3$

$$\begin{cases} (1-\mu)v''(u_{_H})[\triangle\pi - \pi_{_H}\gamma_{_L}(1-\pi_{_L})]\frac{\partial u_{_H}}{\partial\mu} + \mu\pi_{_L}[(1-\pi_{_L})v''(u_{_L}^A) + \pi_{_L}v''(u_{_L}^{NA})]\frac{\partial u_{_L}^A}{\partial\mu} \\ -(1-\mu)v'(u_{_H})\pi_{_H}(1-\pi_{_L})\frac{\partial\gamma_{_L}}{\partial\mu} = \pi_{_L}(1-\pi_{_L})[v'(u_{_L}^{NA}) - v'(u_{_L}^A)] \\ \qquad\qquad +[\triangle\pi - \pi_{_H}\gamma_{_L}(1-\pi_{_L})]v'(u_{_H}) \\ -\pi_{_H}(1-\mu)v''(u_{_H})\left(u_{_L}^A - \underline{u}\right)\frac{\partial u_{_H}}{\partial\mu} - \pi_{_H}(1-\mu)v'(u_{_H})\frac{\partial u_{_L}^A}{\partial\mu} \\ +\mu\pi_{_L}c''(\gamma_{_L})\frac{\partial\gamma_{_L}}{\partial\mu} = -\pi_{_L}c'(\gamma_{_L}) - \pi_{_H}v'(u_{_H})\left(u_{_L}^A - \underline{u}\right) \\ \frac{\partial u_{_H}}{\partial\mu} - \frac{[\triangle\pi - \pi_{_H}\gamma_{_L}(1-\pi_{_L})]}{(1-\pi_{_L})}\frac{\partial u_{_L}^A}{\partial\mu} + \pi_{_H}(u_{_L}^A - \underline{u})\frac{\partial\gamma_{_L}}{\partial\mu} = 0 \end{cases}$$

We obtain the following expressions.

$$\left.\frac{\partial u_{_L}^A}{\partial\mu}\right|_{\underline{u}=\underline{u}^*} = -\frac{(1-\pi_{_L})[\triangle\pi - 2\pi_{_H}\gamma_{_L}(1-\pi_{_L})]\pi_{_L}v'(u_{_H})c}{D} > 0$$

because $\gamma_{_L} < \frac{\tau}{2}$ and

$$\left.\frac{\partial u_{_H}}{\partial\mu}\right|_{\underline{u}=\underline{u}^*} = -\frac{\pi_{_L}v'(u_{_H})c}{2D}\frac{g(\gamma_{_L})}{2(1-\pi_{_L})}$$

where $g(\gamma_{_L}) = 6(\pi_{_H}(1-\pi_{_L})\gamma_{_L})^2 - 7\triangle\pi\pi_{_H}(1-\pi_{_L})\gamma_{_L} + 2\triangle\pi^2$. It is straightforward to verify that $g(\gamma_{_L}) > 0$ when $\gamma_{_L} < \frac{\tau}{2}$. As this is the case when $\underline{u} = \underline{u}^*$ then $\left.\frac{\partial u_{_H}}{\partial\mu}\right|_{\underline{u}=\underline{u}^*} > 0$.

Next we compute $\frac{d\underline{u}^*}{d\mu}$. To do that, we know that

$$\frac{d}{d\mu}\left(\left.\frac{\partial u_{_H}}{\partial\underline{u}}\right|_{\underline{u}=\underline{u}^*}\right) = \frac{1}{D}\left.\frac{dN}{d\mu}\right|_{\underline{u}=\underline{u}^*} = 0$$

When $\underline{u} = \underline{u}^*$, we write N in the following way

$$\begin{aligned} N = \quad & \triangle\pi(1-\mu)v'(u_{_H}(\mu,\underline{u}^*(\mu))) \\ & -2\mu\pi_{_L}[(1-\pi_{_L})v''(u_{_L}^A(\mu,\underline{u}^*(\mu))) \\ & +\pi_{_L}v''(u_{_L}^{NA}(\mu,\underline{u}^*(\mu)))][u_{_L}^A(\mu,\underline{u}^*(\mu)) - \underline{u}^*(\mu)] \end{aligned}$$

and applying the fact that $N = 0$, the second equation in $\mathcal{S}_3$ and rearranging, we obtain

$$\left(\left. \frac{\partial N}{\partial \underline{u}} \right|_{\underline{u}=\underline{u}^*} \right) \frac{d\underline{u}^*}{d\mu} = 2\mu\pi_{_L} H \left. \frac{\partial u_{_L}^A}{\partial \mu} \right|_{\underline{u}=\underline{u}^*}$$

where

$$\begin{aligned} H = \; & \frac{[(1-\pi_{_L})^2 v'''(u_{_L}^A) - (\pi_{_L})^2 v'''(u_{_L}^{NA})]}{1-\pi_{_L}} (u_{_L}^A - \underline{u}) \\ & + \tfrac{3}{2}((1-\pi_{_L}) v''(u_{_L}^A) + \pi_{_L} v''(u_{_L}^{NA})) > 0 \end{aligned}$$

Therefore $\frac{d\underline{u}^*}{d\mu} > 0$.

Knowing that, it is straightforward to verify that

$$(\underline{u}^*(\mu), \mu) \to ((\underline{u})_{Min}, (\widetilde{\mu})_{Min})$$

when μ decreases.

The last result we want to show is the sign of $\frac{d}{d\mu}\left(u_{_L}^A - \underline{u}\right)\Big|_{\underline{u}=\underline{u}^*}$.

$$\begin{aligned} \tfrac{d}{d\mu}\left(u_{_L}^A - \underline{u}\right) &= \tfrac{\partial u_{_L}^A}{\partial \mu} + \tfrac{\partial u_{_L}^A}{\partial \underline{u}} \tfrac{d\underline{u}^*}{d\mu} - \tfrac{d\underline{u}^*}{d\mu} \\ &= \tfrac{\partial u_{_L}^A}{\partial \mu} \left(1 + \left(\tfrac{\partial u_{_L}^A}{\partial \underline{u}} - 1 \right) \tfrac{2\mu\pi_{_L} H}{\left. \frac{\partial N}{\partial \underline{u}} \right|_{\underline{u}=\underline{u}^*}} \right) \end{aligned}$$

As

$$\left. \frac{\partial N}{\partial \underline{u}} \right|_{\underline{u}=\underline{u}^*} < -2\mu\pi_{_L}[(1-\pi_{_L}) v''(u_{_L}^A) + \pi_{_L} v''(u_{_L}^{NA})] \left(\frac{\partial u_{_L}^A}{\partial \underline{u}} - 1 \right)$$

then

$$\left(\frac{\partial u_{_L}^A}{\partial \underline{u}} - 1 \right) \frac{2\mu\pi_{_L} H}{\left. \frac{\partial N}{\partial \underline{u}} \right|_{\underline{u}=\underline{u}^*}} < -\frac{H}{[(1-\pi_{_L}) v''(u_{_L}^A) + \pi_{_L} v''(u_{_L}^{NA})]} < -1$$

So finally $\frac{d}{d\mu}\left(u_{_L}^A - \underline{u}^*\right) < 0$■

Welfare implications of $\underline{u}^*$

We define the social welfare as $\mathcal{W} \equiv \mathop{\mathbb{E}}_{j} \Pi \; + \mu \mathcal{U}_{_L} + (1-\mu)\mathcal{U}_{_H}$. This social criterion is strictly concave in all its arguments. We know that $\mathcal{U}_{_H}$ is an u-shaped function with respect to $\underline{u}$. Moreover, as the feasible set of contracts is reduced when $\underline{u}$ increases, $\mathop{\mathbb{E}}_{j} \Pi$ decreases in that case. So it is immediate to see that $\underline{u}^*$ is not a maximum of $\mathcal{W}$ and by setting $\underline{u}^{\mathcal{W}} < \underline{u}^*$, the expected welfare increases ■