The Geneva Papers on Risk and Insurance Theory, 24: 29–54 (1999)

Optimal Insurance Under Random Auditing

MARIE-CÉCILE FAGART fagart@ensae.fr
University of Rouen, THEMA and LEI

PIERRE PICARD pierre.picard@u-paris10.fr
THEMA (University of Paris X–Nanterre) and CEPREMAP

Abstract

We provide a characterization of an optimal insurance contract (coverage schedule and audit policy) when the monitoring procedure is random. When the policyholder exhibits constant absolute risk aversion, the optimal contract involves a positive indemnity payment with a deductible when the magnitude of damages exceeds a threshold. In such a case, marginal damages are fully covered if the claim is verified. Otherwise, there is an additional deductible that disappears when the damages become infinitely large. Under decreasing absolute risk aversion, providing a positive indemnity payment for small claims with a nonmonotonic coverage schedule may be optimal.

Key words: insurance, auditing, risk aversion

1. Introduction and summary of results

When insurance purchasers have private information about their losses, insurance contracts usually involve monitoring procedures to verify the extent of damages suffered by the policyholders. Insurance contracts should then reach a compromise between two conflicting objectives: sharing the risk between the insurer and the policyholder and minimizing the expected verification cost. This leads to a second-best insurance contract, including a coverage schedule and a monitoring procedure (auditing policy).

A *deterministic auditing policy* specifies whether there is verification or not as a function of the magnitude of damages (Townsend [1979]). Under deterministic auditing, the optimal coverage schedule may be characterized under various assumptions, particularly about what is observed by the insurer and about the ability of the policyholder to manipulate the level of loss (see Bond and Crocker [1997], Gollier [1987], Huberman, Mayers, and Smith [1983], Picard [1999]). In particular, the optimal coverage schedule includes a flat part for small claims for which there is no verification and possibly a deductible when the loss exceeds a threshold.

Under *random auditing*, the insurer decides to verify the claims with a probability that depends on the size of the claim. As shown initially by Townsend [1979], random monitoring procedures can dominate deterministic procedures in a Pareto-sense. Mookherjee and Png [1989] show that, under random auditing, an optimal contract exists if the policyholder exhibits a minimal degree of risk aversion and they establish a number of properties of such an optimal insurance contract. Mookherjee and Png show that the optimal auditing policy is random (whatever the size of the claim) if at equilibrium, in all states of nature, the policyholder can be penalized in case of false claim detected by audit. They also show that it is optimal to reward policyholders in case of audit by paying a higher coverage than when the claim is not verified.

Among the issues left open by Townsend [1979] and Mookherjee and Png [1989], we will consider the following questions:[1]

- Is the audit probability an increasing function of the size of the claim?
- Should this probability be zero for small claims?
- Is a deductible optimal, and, if yes, is this deductible constant when the size of the loss increases?
- Is the policyholder's final wealth monotone in his level of loss?

We are not in position to answer all these questions with full generality. An optimal contract maximizes the expected utility of the policyholder under the participation constraint of the insurer and under incentive compatibility constraints that require the policyholder to always prefer to report his level of loss truthfully. However, it turns out that the standard methods of incentive theory do not allow us to characterize the optimal solution of this problem in a simple way. Technically, there is a continuum of types for the agent, but for some types all the incentive compatibility constraints are tight and for other types none of them is tight. This implies that we cannot consider only local incentive compatibility constraints. In other words, the differential approach of Guesnerie and Laffont [1984] is useless in this model.

Nevertheless, we are able to answer the previous questions when the policyholder has constant absolute risk aversion. We also have partial answers in the case of a nonincreasing absolute risk aversion.

Under constant absolute risk aversion, we are in position to identify the types for which the incentive compatibility constraints are tight and those for which they are not. This will allow us to completely characterize the optimal contract. We then have the following results. The answer to the first two questions above is yes. More precisely, the probability of audit (expressed as a function of the loss) starts from zero at a cutoff point under which no claim is filed and it goes to a limit less than one when the loss goes to infinity. Turning to the third question when a claim is filed (i.e., when the size of the loss exceeds the threshold), then a constant deductible is optimal for verified claims. However, the deductible is larger when the claim is not audited than when it is audited, but the difference decreases and it goes to zero when the size of the loss increases and goes to infinity. In other words, the answer to the third question is that the optimal contract involves a constant deductible to which a vanishing deductible should be added if the claim is not verified. These results imply that the answer to question 4 is no: the policyholder's final wealth decreases when the loss increases but remains under the threshold (no claim is filed in such cases) and the final wealth increases when the loss outweighs the threshold since then a claim is filed with a constant (respectively, decreasing) deductible when the claim is (respectively, is not) verified.

We also show that these answers to the above questions remain valid if the policyholder exhibits nonincreasing absolute risk aversion provided that the smallest claims are not audited in the optimal contract. Under decreasing absolute risk aversion, we show that, in some cases, the optimal insurance contract involves a positive coverage and a positive audit probability for small levels of damages. In such a case the optimal coverage schedule is not monotonic, which confirms a conjecture by Mookherjee and Png [1989] (see note 1).

After presenting the model in Section 2, we state basic properties of the optimal contract in Section 3. Our results here are similar to those of Mookherjee and Png [1989] although our modeling differs in two respects: first, we consider a continuum of possible loss levels, and second, we assume that feasible penalties are upward bounded either because of liquidity constraints or because the size of possible penalties is exogenously determined by law.[2] It is shown in Appendix A that the optimal insurance contract involves random auditing for all claims if feasible penalties are large enough. Section 4 characterizes the optimal insurance contract under random auditing when the policyholder exhibits constant absolute risk aversion. Section 5 provides results for the case of a nonincreasing absolute risk aversion. Section 6 concludes. Most of the proofs are in Appendix B.

2. The model

An insurance buyer owns an initial wealth W, and he faces an uncertain loss with monetary value θ, where θ is a random variable with a support $\Theta = [0, \overline{\theta}]$ and a cumulative distribution $F(\theta)$. We assume that $\theta = 0$ with probability $f(0)$ and that θ is distributed over $(0, \overline{\theta}]$ with a density $f(\theta) = F'(\theta) > 0$. Hence $f(\theta)/1 - f(0)$ is the density of the magnitude of damages over $(0, \overline{x}]$ conditionally on a loss occurring.

The policyholder privately observes his level of loss, and he reports it to the insurer. The latter can verify the damage, but he then incurs an audit cost c. The policyholder experiences a loss $\theta \in \Theta$, and he may choose to file a claim $\hat{\theta} \in \Theta$, which is a message sent to the insurer. The latter commits to audit a claim $\hat{\theta}$ with probability $p(\hat{\theta})$. If the claim is not audited, the payment (net of the insurance premium) from the insurer to the policyholder is denoted $R_N(\hat{\theta})$. In case of audit, the net payment $R_A(\theta, \hat{\theta})$ depends both on the message $\hat{\theta}$ and on the true value of the damage θ.

Let W_f denote the policyholder's final wealth. We have $W_f = W - \theta + R_N(\hat{\theta})$ if the claim is not audited and $W_f = W - \theta + R_A(\theta, \hat{\theta})$ in case of audit. The policyholder is risk-averse. He maximizes the expected utility of his final wealth $EU(W_f)$, where $U(.)$ is a twice differentiable Von Neumann-Morgenstern utility function, with $U' > 0$, $U'' < 0$.

Let $\hat{\theta} = m(\theta)$ be the message sent by the policyholder when he experiences a loss θ. Function $m(.)$ defines the strategy of the policyholder. We have

$$EU(W_f) = \int_\Theta \{[1 - p(m(\theta))]U(W + R_N(m(\theta)) - \theta) \\ + p(m(\theta))U(W + R_A(\theta, m(\theta)) - \theta)\}\, dF(\theta).$$

The insurer is risk neutral. His expected profit $E\Pi$ is

$$E\Pi = -\int_\Theta \{[1 - p(m(\theta))]R_N(m(\theta)) + p(m(\theta))[R_A(\theta, m(\theta)) + c]\}\, dF(\theta),$$

and he is willing to participate if $E\Pi \geq 0$.

Finally, the net payments from the insurer to the policyholder are bounded from below by a maximal penalty B that can be imposed in case of misrepresentation of loss. B may result either from a liquidity constraint or from the fact that penalties are exogenously determined by law.[3]

Hence a *feasible* insurance contract is defined by functions $p(.) : \Theta \to [0, 1]$, $R_N(.) : \Theta \to [-B, +\infty]$ and $R_A(.,.) : \Theta \times \Theta \to [-B, +\infty]$.[4] An optimal insurance contract maximizes $EU(W_f)$ in the set of feasible contracts subject to the insurer's participation constraint $E\Pi \geq 0$, given that $m(.) : \Theta \to \Theta$ is an optimal strategy of the policyholder.

3. General properties of an optimal insurance contracts

This section characterizes basic properties of an optimal contract. First, because of the revelation principle we can restrict attention to incentive compatible contracts where the agent reports his damage truthfully: $m(\theta) \equiv \theta$ is then an optimal strategy for the policyholder. The incentive compatibility conditions are written as

$$\begin{aligned}&[1 - p(\theta)]U(W + R_N(\theta) - \theta) + p(\theta)U(W + R_A(\theta, \theta) - \theta)\\&\quad \geq [1 - p(\hat{\theta})]U(W + R_N(\hat{\theta}) - \theta) + p(\hat{\theta})U(W + R_A(\theta, \hat{\theta}) - \theta) \quad \text{for all } \hat{\theta}, \theta \text{ in } \Theta.\end{aligned}$$

Obviously, it is always optimal to decrease the right-hand side of these incentive compatibility constraints as much as possible. Hence without loss of generality, $R_A(\theta, \hat{\theta}) = -B$ if $\hat{\theta} \neq \theta$ is optimal. In other words, the penalty levied when a report is detected to be false is made as large as possible. Note however that, at equilibrium, nobody lies, so this penalty B is never actually levied.

In what follows, we write $R_A(\theta) \equiv R_A(\theta, \theta)$, and we consider incentive compatible contracts with maximal penalty if the policyholder is detected to have lied. Such contracts are denoted $\{R_N(.), R_A(.), p(.)\}$.

This allows us to characterize the insurance contract in a more usual way. Let

$$P = -\mathrm{Inf}\{R_N(\theta), R_A(\theta), \theta \in \Theta\}$$

and

$$q_N(\theta) = R_N(\theta) + P \geq 0$$
$$q_A(\theta) = R_A(\theta) + P \geq 0.$$

P may be interpreted as the insurance premium paid by the policyholder and $q_N(\theta)$, $q_A(\theta)$ as (nonnegative) indemnity payments. Note that a positive indemnity can be paid in the no-loss state (when $\theta = 0$). In other words, we may have $q_N(0) > 0$ or $q_A(0) > 0$.

The optimal contract maximizes the policyholder's expected utility subject to the incentive compatibility constraints, to the insurer's participation constraint and to feasibility constraints. This may be written as

$$\text{Maximize}\, EU = \int_{\Theta} \{[1 - p(\theta)]U(W - \theta + R_N(\theta)) + p(\theta)U(W - \theta + R_A(\theta))\}\, dF(\theta) \tag{1}$$

with respect to $p(.)$, $R_N(.)$ and $R_A(.)$, subject to

$$\int_{\Theta} \{[1 - p(\theta)]R_N(\theta) + p(\theta)(R_A(\theta) + c)\}\, dF(\theta) \leq 0 \tag{2}$$

$$[1 - p(\theta)]\, U\, (W - \theta + R_N(\theta)) + p(\theta)\, U\, (W - \theta + R_A(\theta)) \geq [1 - p(\hat{\theta})]\, U\, (W - \theta + R_N(\hat{\theta})) + p(\hat{\theta})\, U\, (W - B - \theta) \quad \text{for all } \theta, \hat{\theta} \neq \theta \tag{3}$$

$$R_N(\theta) \geq -B \quad \text{for all } \theta \tag{4}$$

$$R_A(\theta) \geq -B \quad \text{for all } \theta \tag{5}$$

$$0 \leq p(\theta) \leq 1 \quad \text{for all } \theta. \tag{6}$$

This maximization problem will be denoted P_0. Condition (2) is the insurer's participation constraint. Conditions (3) are incentive compatibility constraints: they state that the policyholder is willing to report his level of loss truthfully. (4), (5), and (6) are feasibility conditions.

An optimal solution to P_0 entails some degree of insurance ($R_N(\theta) = 0$, $p(\theta) = 0$ for all θ is *not* an optimal insurance contract) if verifying claims is not too costly (if c is not too large).[5] We assume that this condition is satisfied.

Let

$$v(\theta) \equiv [1 - p(\theta)]U(W + R_N(\theta) - \theta) + p(\theta)U(W + R_A(\theta) - \theta),$$

$v(\theta)$ is the expected utility of a type-θ policyholder. Let us define Assumption **A** as "An optimal contract is such that $v(\theta) > U(W - \theta - B)$ for all θ in Θ." We show in Appendix A that **A** holds when (*ceteris paribus*) B is large enough. If **A** does not hold, then the optimal audit policy is deterministic—that is, $p(\theta) \in \{0, 1\}$ for all θ as in Townsend [1979].[6] In what follows, we always assume that **A** holds.

Proposition 1: *An optimal insurance contract has the following properties*[7]

(i) $p(\theta) < 1$ *for all* θ

(ii) $R_A(\theta) > R_N(\theta)$ *for all* θ *such that* $0 < p(\theta) < 1$

(iii) *If* $p(\hat{\theta}) > 0$ *for some* $\hat{\theta}$ *in* Θ, *then there exists* θ *in* Θ *such that*

$$v(\theta) = [1 - p(\hat{\theta})]U(W - \theta + R_N(\hat{\theta})) + p(\hat{\theta})U(W - \theta - B)$$

(iv) *If* $R_N(\tilde{\theta}) = \text{Min}\{R_N(\theta), \theta \in \Theta\}$, *then* $p(\tilde{\theta}) = 0$. *Furthermore* $p(\theta'') > p(\theta')$ *if* $R_N(\theta'') > R_N(\theta')$.

Proposition 1 is similar to results obtained by Mookherjee and Png [1989].

(i) means that all loss reports are either audited randomly or not audited. Indeed, choosing $p(\theta) = 1$ for some θ would be excessive since the incentive constraints corresponding to the report θ would never be tight, whatever the true level of loss. It would be possible, with an unchanged expected cost, to decrease the audit probability $p(\theta)$ and simultaneously to pay a larger indemnity $R_N(\theta) = R_A(\theta)$, whether the claim θ is audited or not. This transformation would increase the welfare of a policyholder with loss θ, hence the result.

(ii) shows that the insurer should pay a bonus to the policyholder if the report is audited and found to be truthful. Indeed, if we had $R_A(\theta) < R_N(\theta)$ for some θ, the insurer could increase $R_A(\theta)$ and decrease $R_N(\theta)$ in such a way as not to modify the expected cost. He could, for instance, pay $p(\theta)R_A(\theta) + [1 - p(\theta)]R_N(\theta)$ whether the claim is audited or not. This would preserve the incentive compatibility constraints and the type-θ policyholder's expected utility would increase. If such a transformation were possible, the contract would not be optimal. If we had $R_A(\theta) = R_N(\theta)$ and $0 < p(\theta) < 1$, it would be possible to decrease $R_N(\theta)$ and $p(\theta)$ and simultaneously to increase $R_A(\theta)$ in such a way that all incentive constraints still hold, that the expected cost remains constant, and that the expected utility increases.[8]

According to (iii), if a report $\hat{\theta}$ is audited with positive probability, there must exist a level of loss θ such that the policyholder is indifferent between reporting truthfully and reporting $\hat{\theta}$. Suppose otherwise that, whatever the true level of loss, the policyholder always strictly prefers reporting truthfully than reporting $\hat{\theta}$. Then, the insurer can lower $p(\hat{\theta})$ and simultaneously increase $R_N(\hat{\theta})$ in such a way as not to affect the expected cost and as to increase the policyholder's expected utility when the loss is $\hat{\theta}$. The incentive compatibility conditions are preserved if this transformation is small enough.

The first part of (iv) states that any report $\tilde{\theta}$ corresponding to the lowest indemnity payment in the absence of audit will not be audited. The sketch of the proof of this result is as follows. One easily checks that Assumption **A** and (iii) simultaneously imply $R_N(\tilde{\theta}) > -B$. Assume $p(\tilde{\theta}) > 0$. Then, (ii) implies that the policyholder always (strictly) prefers reporting truthfully than reporting $\tilde{\theta}$, which gives $p(\tilde{\theta}) = 0$, hence a contradiction. Intuitively, reporting $\tilde{\theta}$ is always suboptimal when $\theta \neq \tilde{\theta}$ and auditing such a report is costly and useless. The second part of (iv) implies that audit probabilities are increasing in the reported loss if the indemnity payments are increasing. Larger indemnity payments are more tempting for the policyholder and then larger audit probabilities are needed to deter fraudulent claims.

The proof of Proposition 1 shows that any feasible contract that does not satisfy the conditions (i) to (iv) is dominated by another contract. Hence, without loss of generality we may restrict attention to contracts that satisfy this conditions, particularly when one aims at proving the existence of an optimal contract or at characterizing such a contract.

4. Optimal insurance contract under constant absolute risk aversion

The main difficulty in characterizing the optimal contract more completely is to identify the incentive compatibility constraints that are tight at the optimum. This can be done when the policyholder exhibits constant absolute risk aversion—when $U(.)$ is CARA. Let us first state a preliminary lemma.

Lemma 1: *Assume that $U(.)$ is CARA. If $\{R_N(.), R_A(.), p(.)\}$ verifies*

$$R_A(\theta) \geq R_N(\tilde{\theta}) \tag{7}$$

$$U(W + R_N(\tilde{\theta})) \geq [1 - p(\theta)]U(W + R_N(\theta)) + p(\theta)U(W - B) \quad \textit{for all } \theta, \tag{8}$$

where $\tilde{\theta} \in \Theta$ is such that

$$R_N(\tilde{\theta}) = \mathrm{Min}\{R_N(\theta), \theta \in \Theta\}, \tag{9}$$

then all the incentive constraints (3) *are satisfied.*

Proof. When $U(.)$ is CARA, (8) gives

$$\begin{aligned} U(W - \theta + R_N(\tilde{\theta})) \geq\ & [1 - p(\hat{\theta})]U(W - \theta + R_N(\hat{\theta})) \\ & + p(\hat{\theta})U(W - \theta - B) \quad \text{for all } \theta, \hat{\theta} \text{ in } \Theta. \end{aligned} \tag{10}$$

Using (7) and (9) gives

$$\begin{aligned} & [1 - p(\theta)]U(W - \theta + R_N(\theta)) + p(\theta)U(W - \theta + R_A(\theta)) \\ & \quad \geq U(W - \theta + R_N(\tilde{\theta})) \quad \text{for all } \theta \text{ in } \Theta. \end{aligned} \tag{11}$$

(10) and (11) simultaneously give (3). □

Lemma 1 states that at an optimal contract the policyholder is deterred from filing a false claim whatever his true level of loss θ if and only if he is detered from filing a false claim when he has a right to receive the smallest indemnity payment (when $\theta = \tilde{\theta}$). Indeed, when $\theta = \tilde{\theta}$, truthtelling brings in $R_N(\tilde{\theta})$ with certainty, which is less than $R_N(\theta)$ or $R_A(\theta)$ if $\theta \neq \tilde{\theta}$. A false report $\hat{\theta}$ brings in $-B$ with probability $p(\hat{\theta})$ and $R_N(\hat{\theta})$ with probability $1 - p(\tilde{\theta})$. If the propensity to run risks is not affected by wealth effects, which is the case when the policyholder exhibits constant absolute risk aversion, then truthtelling when $\theta = \tilde{\theta}$ implies truthtelling for all other possible levels of loss.

When $U(.)$ is CARA, Proposition 1(ii) and 1(iv) implies that any optimal contract satisfies (7), (8), and (9). Lemma 1 then implies that an optimal contract maximizes $EU(W_f)$ given by (1) with respect to $R_N(.)$, $R_A(.)$, and $p(.)$ subject to Eqs. (2), (4), (6), (7), (8), and (9). Let us call this problem P_1. Hence P_0 and P_1 have the same optimal solutions when $U(.)$ is CARA. Starting from P_1, let us consider a modified problem, called P_2, in which an additional variable K is introduced, with $K \geq -B$, and Eqs. (7)–(9) are replaced respectively by

$$R_A(\theta) \geq K \tag{12}$$

$$U(W + K) \geq [1 - p(\theta)]U(W + R_N(\theta)) + p(\theta)U(W - B) \quad \text{for all } \theta \tag{13}$$

$$R_N(\theta) \geq K \quad \text{for all } \theta. \tag{14}$$

We show in Appendix B (see Lemma 4) that P_1 and P_2 have the same optimal solutions, with $K = R_N(\tilde{\theta}) = \text{Min}\{R_N(\theta), \theta \in \Theta\} > -B$.

Lemma 2 characterizes the optimal audit strategy. The proof of Lemma 2 simply shows that, at the optimum of P_2, (13) is binding for (almost) all θ. Hence, if $R_N(\theta) > K$ at the optimum, then $p(\theta)$ is just large enough to deter the policyholder who is entitled to receive K from putting in a claim corresponding to damages θ.

Lemma 2: *When $U(.)$ is CARA, an optimal audit strategy is such that $p(\theta) = \phi(R_N(\theta), K)$ for all θ with*

$$\phi(R, K) \equiv \frac{U(W+R) - U(W+K)}{U(W+R) - U(W-B)} \in [0, 1) \quad \text{if } R \geq K > -B$$

and $\phi_1' > 0$, $\phi_1'' < 0$, $\phi_2' < 0$.

Proof. At an optimal solution of P_2, (13) is binding for all θ (except perhaps on a zero-measure subset of Θ). Indeed, assume that (13) is not binding for all $\theta \in [a, b] \subset \Theta$, with $a < b$. We then have $p(\theta) > 0$ for all θ in $[a, b]$, otherwise (13) would not hold. A small increase in payments $dR(\theta) \equiv dR_A(\theta) = dR_N(\theta) > 0$ compensated by a decrease in the audit probability $dp\ (\theta) < 0$ such that[9]

$$dR(\theta) = -(R_A(\theta) + c - R_N(\theta))\ dp\ (\theta) \quad \text{for all } \theta \text{ in } [a, b]$$

is feasible in P_2. In particular, it does not modify the expected cost incurred by the insurer. Such a change improves the expected utility of the policyholder, hence a contradiction. □

Function $\phi(., K)$ is depicted in figure 1 when $-B < K < 0$. When the insurance policy assigns a net income K with certainty to the policyholder, $\phi(R, K)$ is the audit probability that deters him from misrepresenting his loss in the hope of receiving R if he is not detected. When the policyholder exhibits constant absolute risk aversion, this probability does not depend on the true level of damages. ϕ is increasing and concave with respect to R and

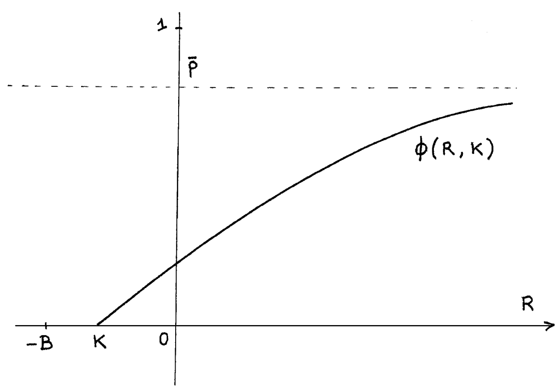

Figure 1. The audit probability function.

(because a CARA utility function is upward bounded) ϕ goes to a limit $\bar{p}$ less than one, when R goes to infinity. We are now in position to prove that an optimal insurance contract exists and to characterize it.

Proposition 2: *When $U(.)$ is CARA, there exists an optimal contract. It has the following properties*:

(i) $R_A(\theta) = M + \theta$ *for all* θ *such that* $R_N(\theta) > K$ *where* M *is a constant such that* $M < K < 0$.

(ii) *There exists a threshold* $\theta^* > 0$ *such that*

$$R_N(\theta) = K \quad \text{if } \theta \leq \theta^*$$
$$R_N(\theta) > K \quad \text{if } \theta > \theta^*.$$

(iii) $R_N(\theta) = R_A(\theta) - \eta(\theta) = M + \theta - \eta(\theta)$ *if* $\theta > \theta^*$ *where* $\eta(\theta)$ *is a continuous function with* $\eta(\theta) > 0$, $\eta(\theta^*) = M + \theta^* - K > 0$, $\eta'(\theta) < 0$, $\eta(\theta) \to 0$ *when* $\theta \to +\infty$.

(iv) *The optimal audit strategy is such that* $p(\theta) = 0$ *if* $0 < \theta \leq \theta^*$ *and* $p(\theta) > 0$, $p'(\theta) > 0$ *if* $\theta > \theta^*$.

Existence of an optimal contract follows from the fact that a CARA utility function is upward bounded: the policyholder cannot be induced to tell the truth in contracts where claims are audited with very low probability and where very large indemnity payments are provided for claims that are verified to be truthful (see Border and Sobel [1987] and Mookherjee and Png [1989]).

Proposition 2 provides a full characterization of an optimal insurance policy and of the corresponding audit strategy. The insurer's participation constraint implies $K < 0$. Hence $P = -K$ may be interpreted as the insurance premium paid by the policyholder and $q_N(\theta) \equiv R_N(\theta) - K$ and $q_A(\theta) \equiv R_A(\theta) - K$ correspond to the indemnity payment. The optimal coverage schedule and the audit strategy are depicted in figure 2. There exists a positive threshold θ^* below which no claim is filed (that is, $q_N(\theta) = 0$ and $p(\theta) = 0$ if $\theta \leq \theta^*$) that corroborates a conjecture by Townsend [1979] (see note 1). When the size of the loss exceeds θ^*, the policyholder receives a positive insurance coverage with a deductible, and the claim is verified with positive probability. If the claim is verified, the deductible is constant and equal to $\theta - q_A(\theta) = K - M = m > 0$. If the claim is not verified, the deductible is larger than m: it is equal to $m + \eta(\theta)$. The additional deductible $\eta(\theta)$ is decreasing, and it goes to zero when θ it goes to infinity (when $\bar{\theta} = +\infty$): The probability of audit increases when the size of the claim increases (that is, $p'(\theta) > 0$) and $p(\theta) \to \bar{p} < 1$ when $\theta \to +\infty$. In other words, marginal damages are fully covered in case of audit, and, in the other case, there is an additional deductible that disappears when the damages become infinitely large.

The intuition for these results is as follows. Consider a loss level θ larger than θ^*. Let $dR_N(\theta)$ be an infinitesimal increase in $R_N(\theta)$. Such an increase affects the insurer's cost in two ways: directly because of the increase in the payment to the policyholder and indirectly through its effect on the audit probability $p(\theta)$. Lemma 2 shows that the increase $dR_N(\theta)$ should be accompanied by an increase $dp(\theta) = \phi_1'(R_N(\theta), K)\, dR_N(\theta)$ for incentive compatibility to be maintained. Hence the total effect on the insurer's expected cost

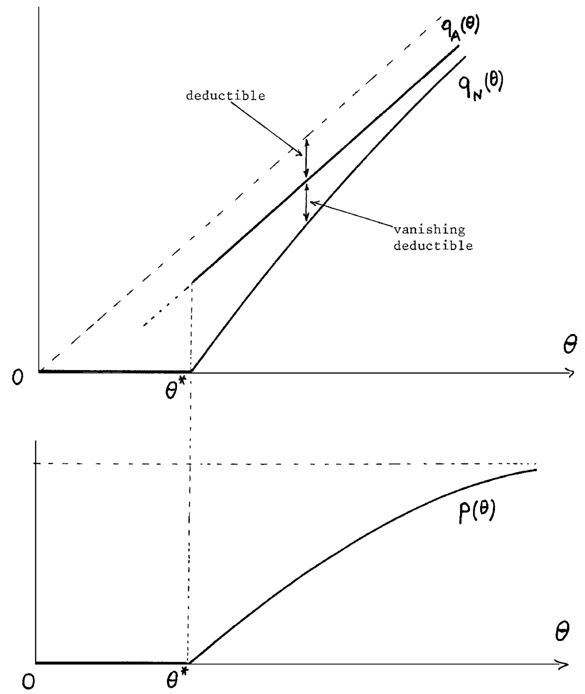

Figure 2. The optimal insurance policy and the audit strategy.

(conditionally on the occurrence of a loss of size θ) is $dC = [1 + c\phi_1'(R_N(\theta), K)]\, dR_N(\theta)$. On the contrary, an increase in $R_A(\theta)$ does not affect the audit strategy, and we have $dC = dR_A(\theta)$ in that case. In other words, the incremental cost associated with one more dollar paid in state θ is larger when the additional payment affects the claims that are not audited than when it affects the claims that are audited. Consequently, it is optimal to choose $R_A(\theta) > R_N(\theta)$, which is another way to state Proposition 1(ii) when $U(.)$ is CARA. The fact that $\phi_1'' < 0$ and $R_N'(\theta) > 0$ implies that the incremental audit cost $c\phi_1'(R_N(\theta), K)$ associated with a one dollar increase in $R_N(\theta)$ is decreasing in θ and that this incremental audit cost goes to zero when θ goes to infinity. This explains why the additional deductible $\eta(\theta)$ is decreasing and why it disappears when θ is large.

Under the assumptions of Proposition 2, $v(\theta)$ is continuous over Θ. It reachs a minimum at $\theta = \theta^*$ and $v(\theta) \to U(W + M)$ when $\theta \to +\infty$ (see figure 3). When $\theta < \theta^*$, no insurance payment is made and $v(\theta)$ is decreasing. When $\theta > \theta^*$, the policyholder receives an indemnity payment. Since the additional deductible $\eta(\theta)$ is decreasing, $v(\theta)$ is increasing.

5. Further results

Matters are much more intricate when the assumption of a constant absolute risk aversion is no more made. In that case, the variations in the level of loss entail a wealth effect

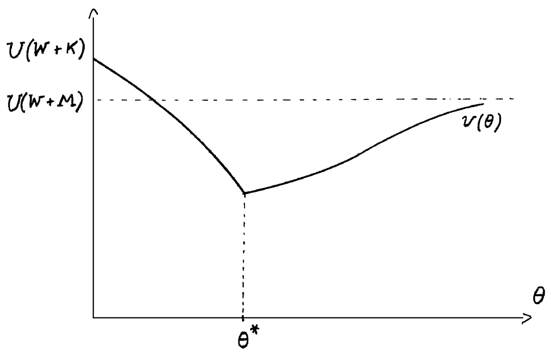

Figure 3. The expected utility of the policyholder.

that prevents us to identify which incentive compatibility constraints will be tight at the optimum. Nevertheless, a partial characterization will be provided when the policyholder exhibits nonincreasing absolute risk aversion. Two effects may be distinguished in such a case. First of all, because of the risk-sharing objective, larger damages should lead to larger indemnity payments. If we only focus on this objective, the net insurance payment should be minimal when $\theta = 0$, which corresponds to the optimal solution derived in the previous section with $q_N(\theta) = 0$ and $p(\theta) = 0$ when θ is small. However, when the policyholder's absolute risk aversion is decreasing with respect to wealth, then it may be optimal to choose an insurance contract that provides a positive coverage (and correlatively a positive probability of audit) for small losses. The intuition for this result is that, under decreasing absolute risk aversion, it is particularly difficult to incite a type-θ policyholder to tell the truth when θ is small because of a wealth effect. Indeed, when $U(.)$ is DARA, a policyholder who has experienced a small level of loss (or no loss at all) inclines to gamble on the audit probability more than a policyholder who has suffered larger damages. In such a setting, paying a positive coverage and incurring verification costs when losses are small is a way to mitigate the intensity of the incentive constraints. It turns out that, under some circumstances, this wealth effect may dominate the risk sharing effect. A nonmonotonic indemnity schedule with positive coverage for small losses will be optimal in such a case.

Let us first state a preliminary lemma that extends Lemma 1 to the case of a nonincreasing absolute risk aversion (NIARA).

Lemma 3: *Assume that $U(.)$ is NIARA. Then any optimal contract maximizes $EU(W_f)$ with respect to $R_N(.)$, $R_A(.)$ and $p(.)$ and $\theta_0 \in \Theta$ subject to* (2), (4), (6), *and*

$$R_A(\theta) \geq R_N(\theta_0) \tag{15}$$

$$\left.\begin{aligned} U(W - \theta_0 + R_N(\theta_0)) \geq [1 - p(\hat{\theta})]\, U\,(W - \theta_0 + R_N(\hat{\theta})) \\ + p(\hat{\theta})\, U\,(W - \theta_0 - B) \quad \textit{for all } \hat{\theta} \end{aligned}\right\} \tag{16}$$

$$\left.\begin{aligned}&[1-p(\theta)]\,U\,(W-\theta+R_N(\theta))+p(\theta)\,U\,(W-\theta+R_A(\theta))\\&\quad\geq[1-p(\hat{\theta})]\,U\,(W-\theta+R_N(\hat{\theta}))+p(\hat{\theta})\,U\,(W-\theta-B)\\&\qquad\qquad\qquad\qquad\qquad\textit{for all } \theta<\theta_0 \quad \textit{and all } \hat{\theta}\neq\theta\end{aligned}\right\} \tag{17}$$

$$R_N(\theta) \geq R_N(\theta_0) \quad \textit{for all } \theta \tag{18}$$

$$R_N(\theta) > R_N(\theta_0) \quad \textit{if } \theta < \theta_0. \tag{19}$$

θ_0 given by (18) and (19) is the smallest level of loss for which the net transfer to the policyholder is minimal: $-R_N(\theta_0)$ may be interpreted as the insurance premium. We know from Proposition 1(iv) that a claim $\hat{\theta}=\theta_0$ will not be audited. We also know from Proposition 1 that the optimal contract verifies the conditions that are listed in Lemma 3. The missing constraints correspond to the incentive compatibility constraints in the case where the magnitude of damages is larger than θ_0. Lemma 3 states that these constraints are superfluous. The intuition for this result is straightforward: if a type-θ_0 policyholder is not willing to gamble on the insurer's audit strategy by announcing a claim $\hat{\theta}\neq\theta_0$, then a policyholder who has experienced a loss larger than θ_0 will also tell the truth for two reasons: first, because truthtelling allows him to get a payment that is larger than (or equal to) that received by the type-θ_0 policyholder and second because he exhibits an absolute risk aversion that is larger than (or equal to) that of the type-θ_0 policyholder.

Proposition 3: *Assume that $U(.)$ is NIARA. Let θ_0 be defined by* (18) *and* (19)*. An optimal contract has the following properties*:
(i) *There exists a constant M such that $R_A(\theta)=M+\theta$ if $\theta>\theta_0$ and $p(\theta)>0$.*
(ii) *If $\theta_0=0$, the optimal contract has the same characterization as in Proposition* 2.

Proposition 3 provides a partial characterization of an optimal insurance policy in the case where the policyholder exhibits nonincreasing absolute risk aversion. The incentive compatibility constraints of a type-θ policyholder are not binding when θ is larger than θ_0 and $p(\theta)$ is strictly positive—when $\theta>\theta_0$ and $R_N(\theta)>R_N(\theta_0)-$. For these levels of θ, variations in the indemnity payment $R_A(\theta)$ would entail no incentive effect and, as stated in Proposition 3(i) the optimal insurance policy provides full insurance at the margin in such a case. On the contrary, some incentive compatibility constraints may be binding if θ is smaller than θ_0 and, for that very reason, it is optimal to increase $R_N(\theta)$ over $R_N(\theta_0)$ in $[0,\theta_0)$. When $\theta_0=0$, the only incentive compatibility constraints that are binding are those of a policyholder who has not experienced any loss and the characterization given in Proposition 2 remains valid. However, when the policyholder exhibits decreasing absolute risk aversion (DARA), the risk sharing objective conflicts with the incentive compatibility constraint. Indeed, in the neighborhood of $\theta=0$, increasing $R_N(\theta)$ over $R_N(\theta_0)$ allows the insurer to relax the incentive compatibility constraints for the type-θ policyholder. On the contrary, the risk sharing objective should lead the insurer to offer a coverage schedule that is increasing in the level of loss. As stated in Proposition 4, it turns out that in some cases the tradeoff may tip in favor of the incentive compatibility concern.

Proposition 4: *Let $F(\theta, z)$ be the cumulative probability distribution of θ over Θ, continuously parametrized by $z \in R_+$, with $f(0, z) = F(0, z) =$ Prob$(\theta = 0 \mid z)$ and $f(\theta, z) = F_1'(\theta, z)$ for $\theta > 0$. Assume that $f(0, 0) = f(0_+, 0) = 0$ and that $f(\theta, z) > 0$ for all θ in Θ if $z > 0$. Assume that $U(.)$ is DARA.*

Let $\theta_0(z)$ be the value of θ_0 defined in Lemma 3 when θ is distributed according to $F(\theta, z)$. Then there exists $\hat{z} > 0$ such that $\theta_0(\hat{z}) > 0$.

Proposition 4 shows that providing positive coverage for small losses is optimal when the policyholder exhibits decreasing absolute risk aversion if the probability of small losses is small enough. Indeed, assume that $f(0) = f(0_+) = 0$. In such a case, providing a small positive coverage in the neighborhood of $\theta = 0$ entails no first-order risk-sharing effect but this positive coverage allows us to relax the incentive compatibility constraints for the policyholders who have experienced very small losses. Consequently, a less intense audit strategy may be chosen: one may lower the audit probability that corresponds to claims $\hat{\theta}$ for which the incentive compatibility conditions are binding when θ is in the neighborhood of $\theta = 0$. This change will entail a positive first-order effect on the policyholder's expected utility. By a continuity argument, this characterization still holds if the policyholder experiences small or zero losses with positive (but small) probability.

6. Conclusion

In this article, we have approached the normative analysis of optimal insurance contracts under costly state verification initiated by Townsend [1979]. We have focused on questions tackled but not completely settled by Mookherjee and Png [1989]—namely, what should be the shape of the coverage schedule and the audit strategy when there is random auditing?

The characterization obtained under constant absolute risk aversion is rather intuitive. There is no insurance coverage when the magnitude of damages is less than a threshold. When the damages exceed the threshold, the policyholder receives a positive indemnity payment with a deductible. This deductible is constant when the claim is verified. Otherwise, there is an additional deductible that reflects the effect of insurance payments on verification costs and that disappears when the claims are very large. Furthermore, the larger the size of the claim, the larger the probability of audit should be.

Under decreasing absolute risk aversion, because of a wealth effect it may be optimal to provide a positive coverage for small claims to mitigate the intensity of the incentive compatibility constraints. In such a case, the optimal coverage schedule may be nonmonotonic, which confirms a conjecture by Mookherjee and Png [1989]. This last result may be considered only as a theoretical curiosity since, in practice, coverage schedules are almost always monotonic. However, it also provides another case for offering contracts with no-claim bonuses besides the well-known interpretations of experience rating in term of effort incentives or informational learning. Of course, the empirical relevance of this argument for no-claim (or small claim) bonuses remains an open issue. If the optimal coverage schedule is monotonic—which is likely if the probability of zero loss or of small losses is large enough—then the characterization derived in the case of a constant absolute risk aversion still holds when the policyholder exhibits decreasing absolute risk aversion.

Appendix A

1. Let us show that *the optimal audit policy is deterministic when assumption* **A** does not hold. In what follows, the index $*$ refers to an optimal solution to P_0. Assume that **A** does not hold for that optimal solution. Let $\theta_0 \in \Theta$ such that $v^*(\theta_0) = U(W - \theta_0 - B)$. If $R_N^*(\theta_1) > -B$ for some θ_1 in Θ, then $p^*(\theta_1) = 1$ for otherwise (3) would not hold for $\theta = \theta_0$ and $\hat{\theta} = \theta_1$. Hence, without loss of generality, we may assume that $R_N^*(\theta) = -B$ for all θ. $R_A^*(.)$, $p^*(.)$ maximize

$$EU = \int_\Theta \{[1 - p(\theta)]U(W - \theta - B)\, dF(\theta) + p(\theta)U(W - \theta + R_A(\theta))\}\, dF(\theta)$$

with respect to $R_A(.)$, $p(.)$, subject to

$$\int_\Theta \{p(\theta)[R_A(\theta) + c] - [1 - p(\theta)]B\}\, dF(\theta) \leq 0$$
$$R_A(\theta) \geq -B \text{ and } 0 \leq p(\theta) \leq 1 \quad \text{for all } \theta.$$

Note that the optimal solution to this problem verifies the incentive compatibility conditions since $R_N^*(\theta) = -B$ for all θ. Furthermore, we have $R_A^*(\theta) > -B$ for all θ such that $p^*(\theta) > 0$ for otherwise $p(\theta) = 0$ would lead to the same expected utility at a lower cost.

Let λ be the Kuhn-Tucker multiplier associated with the insurer's participation constraint in this problem. The first-order optimality conditions are

$$p(\theta)[U'(W - \theta + R_A(\theta)) - \lambda] = 0 \tag{20}$$

and

$$\left.\begin{aligned} \Gamma(\theta) \ &\leq 0 \quad \text{if } p(\theta) = 0 \\ &= 0 \quad \text{if } 0 < p(\theta) < 1 \\ &\geq 0 \quad \text{if } p(\theta) = 1 \end{aligned}\right\} \tag{21}$$

for all θ, where

$$\Gamma(\theta) \equiv U(W - \theta + R_A(\theta)) - U(W - \theta - B) - \lambda[R_A(\theta) + c + B].$$

Using (20), we may write $R_A(\theta) = M + \theta$ for all θ such that $p(\theta) > 0$, where M is a constant such that $M + \theta > -B$ if $p(\theta) > 0$. We have

$$\Gamma'(\theta) = U'(W - \theta - B) - U'(W + M) > 0.$$

Let $\hat{\theta}$ such that $\Gamma(\hat{\theta}) = 0$ with $\hat{\theta} > -B - M$. If $\theta > \hat{\theta}$, we have $\Gamma(\theta) > 0$ which implies $p(\theta) = 1$. If $\theta < \hat{\theta}$, then $\Gamma(\theta) < 0$ and $p(\theta) = 0$. Furthermore, we have $\Gamma(0) < 0$ which implies $\hat{\theta} > 0$. Choosing $p(\theta) = 1$ if $\theta > \hat{\theta}$ and $p(\theta) = 0$ if $\theta \leq \hat{\theta}$ is optimal.

2. Let us show that *Assumption* **A** *holds when B is large enough*. Let us first characterize the optimal contract with deterministic auditing when $B = +\infty$. Under deterministic auditing an incentive compatible contract entails a flat coverage schedule in the no-verification regime, and full insurance at the margin in the verification regime. Furthermore, the largest claims should be verified. Hence we have[10]

$$\begin{aligned} R_N(\theta) &= K \quad \text{and} \quad p(\theta) = 0 \quad \text{if } \theta \leq \hat{\theta} \\ R_A(\theta) &= M + \theta \quad \text{and} \quad p(\theta) = 1 \quad \text{if } \hat{\theta} < \theta \leq \bar{\theta}. \end{aligned}$$

The optimal insurance contract maximizes

$$EU = \int_0^{\hat{\theta}} U(W - \theta + K)\, dF(\theta) + U(W + M)[1 - F(\hat{\theta})]$$

with respect to K, M, and $\hat{\theta} \in [0, \bar{\theta}]$, subject to the insurer's participation constraint

$$KF(\hat{\theta}) + \int_{\hat{\theta}}^{\bar{\theta}} (M + \theta + c)\, dF(\theta) \leq 0$$

and to the incentive compatibility constraint $M + \hat{\theta} \geq K$. Let $K = \bar{K}$ at the optimum under deterministic auditing. This solution remains optimal (under deterministic auditing) when $B > -\bar{K}$.

Assume that Assumption **A** does not hold at the optimum of P_0 and that $B > -\bar{K}$. Then the optimal insurance contract involves deterministic auditing, which implies $R_N(\theta) \geq \bar{K}$ and $R_A(\theta) \geq \bar{K}$ for all θ. Consequently, we have $v(\theta) \geq U(W - \theta + \bar{K}) > U(W - \theta - B)$ for all θ. Hence **A** is satisfied, which is a contradiction.

Appendix B

Proof of Proposition 1

Proof of (i). Assume that, for an optimal contract C, we have $p(\theta) = 1$ for all θ in $[a, b] \subset \Theta$, with $a < b$. Let $\bar{C} = \{\bar{R}_N(.), \bar{R}_A(.), \bar{p}(.)\}$: $\bar{C}$ coincides with C if $\theta \notin [a, b]$, and $\bar{R}_N(\theta) = \bar{R}_A(\theta) = R_A(\theta) + c\sigma$ and $\bar{p}(\theta) = 1 - \sigma$ if $\theta \in [a, b]$, with $\sigma > 0$. $\bar{C}$ is incentive compatible for σ small enough because of Assumption **A**. Furthermore $\bar{C}$ induces the same expected cost as C. $\bar{C}$ provides a higher (an unchanged) expected utility than C if θ is (is not) in $[a, b]$, which contradicts the fact that C is optimal.

Proof of (ii). Assume first that, for an optimal contract C, we have $0 < p(\theta) < 1$ and $R_A(\theta) < R_N(\theta)$ for all θ in $[a, b] \subset \Theta$, with $a < b$.

Let $\bar{C} = \{\bar{R}_N(.), \bar{R}_A(.), \bar{p}(.)\}$ be another contract that coincides with C if $\theta \notin [a, b]$ and $\bar{R}_N(\theta) = \bar{R}_A(\theta) = [1 - p(\theta)]R_N(\theta) + p(\theta)R_A(\theta)$ and $\bar{p}(\theta) = p(\theta)$ if $\theta \in [a, b]$.

Let $\bar{v}(\theta) \equiv [1 - \bar{p}(\theta)]U(W - \theta + \bar{R}_N(\theta)) + \bar{p}(\theta)U(W - \theta + \bar{R}_A(\theta))$.

We have $\bar{v}(\theta) = v(\theta)$ if $\theta \notin [a, b]$ and $\bar{v}(\theta) > v(\theta)$ if $\theta \in [a, b]$ and $\bar{R}_N(\theta) \leq R_N(\theta)$. Hence, using (3) gives

$$\begin{aligned}\bar{v}(\theta) &\geq v(\theta) \geq [1 - p(\hat{\theta})]\, U\,(W - \theta + R_N(\hat{\theta})) + p(\hat{\theta})\, U\,(W - \theta - B) \\ &\geq [1 - p(\hat{\theta})]\, U\,(W - \theta + \bar{R}_N(\hat{\theta})) + p(\hat{\theta})\, U\,(W - \theta - B) \quad \text{for all } \theta \end{aligned} \tag{22}$$

Equation (22) implies that $\bar{C}$ is incentive compatible. Furthermore, $\bar{C}$ induces the same expected cost as C and it provides a higher expected utility to the policyholder. Hence a contradiction.

Assume now that, for an optimal contract C, $0 < p(\theta) < 1$ and $R_A(\theta) = R_N(\theta) > -B$ for all θ in $[a, b] \subset \Theta$, with $a < b$. Consider infinitesimal variation $dR_A(\theta) > 0$, $dR_N(\theta) < 0$, $dp(\theta) < 0$ with $dp(\theta) = h(\theta)\, dR_N(\theta)$ where

$$h(\theta) = \text{Min}\left\{ \frac{[1 - p(\theta)]U'(W - \theta' + R_N(\theta))}{U(W - \theta' + R_N(\theta)) - U(W - \theta' - B)}, \theta' \in \Theta \right\} > 0,$$

which reduces the expected utility of a policyholder who falsely reports that his loss is θ, whatever his true level of loss. Furthermore, the expected cost that corresponds to a type-θ policyholder remains constant—that is,

$$p(\theta)\, dR_A(\theta) + (1 - p(\theta))\, dR_N(\theta) + c\, dp(\theta) = 0,$$

which gives

$$dv(\theta) = -c\, U'(W - \theta + R_N(\theta))\, dp(\theta) > 0.$$

Hence, the expected utility of a type-θ policyholder increases. This transformation may be performed for all θ in $[a, b]$, without modifying the payment and the audit probability when $\theta \notin [a, b]$, which contradicts the fact that C is optimal.

Proof of (iii). Assume that, for an optimal contract C, we have $p(\hat{\theta}) > 0$ and

$$v(\theta) > [1 - p(\hat{\theta})]\, U\,(W - \theta + R_N(\hat{\theta})) + p(\hat{\theta})\, U\,(W - \theta - B)$$

for all $\hat{\theta}$ in $[a, b] \subset \Theta$ with $a < b$ and all $\theta \in \Theta$. Let $\bar{C} = \{\bar{R}_N(.), \bar{R}_A(.), \bar{p}(.)\}$ be another contract that coincides with C if $\theta \notin [a, b]$ and such that

$$\bar{R}_N(\theta) = R_N(\theta) + \varepsilon(\theta), \quad \bar{R}_A(\theta) = R_A(\theta), \quad \bar{p}(\theta) = p(\theta) - \sigma(\theta) \quad \text{if } \theta \in [a, b]$$

with $p(\theta) > \sigma(\theta) > 0$ for all θ in $[a, b]$ and $\varepsilon(\theta) = \eta(\sigma(\theta), \theta)$, with

$$\eta(\sigma, \theta) \equiv \frac{\sigma[R_A(\theta) + c - R_N(\theta)]}{1 - p(\theta) + \sigma}.$$

If $\sigma(\theta)$ is small enough for all θ in $[a, b]$, then $\bar{C}$ is incentive compatible. Furthermore, we have for all θ in $[a, b]$

$$\bar{p}(\theta)[\bar{R}_A(\theta) + c] + [1 - \bar{p}(\theta)]\bar{R}_N(\theta) = p(\theta)[R_A(\theta) + c] + [1 - p(\theta)]R_N(\theta).$$

Hence C and $\bar{C}$ induce the same expected cost. Let $\phi(\sigma(\theta), \theta)$ denote the expected utility of a type-θ policyholder under $\bar{C}$, with

$$\begin{aligned}\phi(\sigma, \theta) = {} & [1 - p(\theta) + \sigma]\, U\,(W - \theta + R_N(\theta) + \eta(\sigma, \theta)) \\ & + [p(\theta) - \sigma]\, U\,(W - \theta + R_A(\theta)).\end{aligned}$$

We have

$$\begin{aligned}\frac{\partial \phi}{\partial \sigma_{|\sigma=0}} = {} & U(W - \theta + R_N(\theta)) - U(W - \theta + R_A(\theta)) \\ & + [R_A(\theta) + c - R_N(\theta)]U'(W - \theta + R_N(\theta)) > 0.\end{aligned}$$

Hence, we can choose a function $\sigma(\theta)$ that is positive and continuous over $[a, b]$ such that the expected utility of any type-θ policyholder increases if $\theta \in [a, b]$, which contradicts the fact that C is an optimal contract.

Proof of (iv). Let $\tilde{\theta} \in \Theta$ such that $R_N(\tilde{\theta}) = \mathrm{Min}\{R_N(\theta), \theta \in \Theta\}$. Assume that $p(\tilde{\theta}) > 0$ which implies $R_N(\tilde{\theta}) > -B$ because of Assumption **A**. Using (ii) gives

$$\begin{aligned}v(\theta) & \geq U(W - \theta + R_N(\tilde{\theta})) \\ & > [1 - p(\tilde{\theta})]U(W - \theta + R_N(\tilde{\theta})) + p(\tilde{\theta})U(W - \theta - B)\end{aligned}$$

for all θ, which implies $p(\tilde{\theta}) = 0$, hence a contradiction. Hence $p(\tilde{\theta}) = 0$. Let

$$\psi(R, \theta) \equiv \frac{U(W + R - \theta) - v(\theta)}{U(W + R - \theta) - U(W - B - \theta)} \quad \text{for } R > -B.$$

Assume $R_N(\theta'') > R_N(\theta') > B$ for θ'', θ' in Θ. We then have $p(\theta'') > 0$, otherwise the type-$\tilde{\theta}$ policyholder would (strictly) prefer to report θ'' rather than $\tilde{\theta}$. Hence, from (iii), there exists θ_0 and θ_1 (if $p(\theta') > 0$) in Θ such that

$$\begin{aligned}p(\theta'') &= \psi(R_N(\theta''), \theta_0) \geq \psi(R_N(\theta''), \theta) \quad \text{for all } \theta \\ p(\theta') &= \psi(R_N(\theta'), \theta_1) \geq \psi(R_N(\theta'), \theta) \quad \text{for all } \theta.\end{aligned}$$

$\psi(R, \theta)$ is increasing with respect to R, which gives

$$p(\theta'') = \psi(R_N(\theta''), \theta_0) \geq \psi(R_N(\theta''), \theta_1) > \psi(R_N(\theta'), \theta_1) = p(\theta').$$

If $p(\theta') = 0$, we also have $p(\theta'') > p(\theta')$. When $R_N(\theta') = -B$, we have $p(\theta') = 0 < p(\theta'')$, which completes the proof. □

Lemma 4: *When U is CARA, an optimal contract $\{R_N(.), R_A(.), p(.), K\}$ is an optimal solution to problem P_2. Furthermore, at the optimum, there exists $\tilde{\theta}$ in Θ such that $K = R_N(\tilde{\theta}) = \text{Min}\{R_N(\theta), \theta \in \Theta\} > -B$.*

Proof. In problem P_2, $EU(W_f)$ is maximized with respect to $R_N(.)$, $R_A(.)$, $p(.)$ and K subject to (2), (6), (12), (13), (14), and $K \geq -B$. At an optimal solution of P_2, EU is at least as high as in P_1 since $K = \text{Min}\{R_N(\theta), \theta \in \Theta\}$ is a possible choice in P_2.

In P_2, the first-order optimality conditions with respect to $R_N(\theta)$ and $R_A(\theta)$ imply $R_A(\hat{\theta}) \geq R_N(\hat{\theta})$. Assume $R_N(\theta) > K$ for all θ—and thus $R_A(\theta) > K$ for all θ—at an optimal solution of P_2. Then K can be slightly increased without violating constraints (12) and (14). Hence (13) is not binding for all θ. Given $p(.)$ the first-order optimality conditions corresponding to the maximization of EU subject to (2) give $R_A(\theta) = R_N(\theta)$ for all θ, which implies that $p(\theta) = 0$ for all θ is optimal. Equation (13) would not hold, hence a contradiction.

Thus, at an optimal solution of P_2, there exists $\tilde{\theta} \geq 0$ such that

$$K = R_N(\tilde{\theta}) = \text{Min}\{R_N(\theta), \theta \in \Theta\}.$$

We deduce that P_1 and P_2 have the same optimal solution. Under assumption **A**, we have $K > -B$ at the optimum. □

Proof of Proposition 2

We first prove that an optimal contract exists. Assume that $U(.)$ is CARA. We know from Lemmas 2 and 4 that any optimal contract maximizes

$$EU = \int_\Theta \{[1 - \phi(R_N(\theta), K)]\, U\,(W + R_N(\theta) - \theta) \\ + \phi(R_N(\theta), K)\, U\,(W + R_A(\theta) - \theta)\}\, dF(\theta) \tag{23}$$

with respect to $R_N(.) R_A(.)$ and $K \geq -B$ subject to

$$\int_\Theta \{[1 - \phi(R_N(\theta), K)] R_N(\theta) + \phi(R_N(\theta), K)[R_A(\theta) + c]\}\, dF(\theta) \leq 0 \tag{24}$$

$$R_N(\theta) \geq K \quad \text{for all } \theta \tag{25}$$

$$R_A(\theta) \geq K \quad \text{for all } \theta. \tag{26}$$

A straightforward extension of Lemmas 1, 2, and 4 shows that, if an optimal contract does not exist, then there exists a sequence of contracts that satisfy Eqs. (24)–(26) such that the corresponding level of expected utility goes to the upper bound of the level of feasible expected utility. Hence, to prove the existence of an optimal contract we have to prove that the above maximization problem has an optimal solution.

The proof is in two steps. In Step 1, we consider K as a fixed parameter (with $-B \leq K \leq 0$), and we characterize an optimal contract conditionally on K. In Step 2, we show that the expected utility that corresponds to this optimal contract (denoted by $\Omega(K)$ hereafter) is continuous w.r.t. $K \in [-B, 0]$, which proves that $\Omega(K)$ reaches a maximum in $[-B, 0]$.

Step 1. Consider K as fixed parameter, with $K < 0$. Let

$$\begin{aligned}\mathcal{L}(R_N, R_A, \lambda, \theta) = \{&[1-\phi(R_N, K)]\,U\,(W + R_N - \theta)\\ &+\phi(R_N, K)\,U\,(W + R_A - \theta)\}f(\theta)\\ &-\lambda\{[1-\phi(R_N, K)]R_N + \phi(R_N, K)[R_A + c]\}f(\theta)\end{aligned}$$

be the Lagrangean, where λ is a Kuhn-Tucker multiplier. If there exist $\lambda \geq 0$ and $\{R_N(.), R_A(.)\}$ such that, for all θ, $(R_N(\theta), R_A(\theta))$ maximizes $\mathcal{L}(R_N, R_A, \lambda, \theta)$ subject to (25), (26) and (24) holds, then $\{R_N(.), R_A(.)\}$ is an optimal contract (conditionally on K) (see Hestenes [1980], Theorem 5.1).

Assume $\lambda > 0$. Let $\tilde{R}_A(\theta, \lambda, K) = \text{Sup}\{\tilde{M}(\lambda) + \theta, K\}$, where $\tilde{M}(\lambda)$ is given by $U'(W + \tilde{M}) = \lambda$. Note that $\tilde{M}$ always exists when $U(.)$ is CARA. $\tilde{R}_A(\theta, \lambda)$ maximizes $\mathcal{L}(R_N, R_A, \lambda, \theta)$ with respect to R_A, subject to (26). Furthermore, $\tilde{R}_A(\theta, \lambda)$ is continuous with respect to θ and λ.

Let us consider two possible cases. Assume first that $\tilde{R}_A(\theta, \lambda, K) = \tilde{M}(\lambda) + \theta$. Let

$$\tilde{H}(R_N, \theta, \lambda) = \frac{\mathcal{L}_1'(R_N, \tilde{M}(\lambda) + \theta, \lambda, \theta)}{\phi_1'(R_N, K)f(\theta)}$$

—that is,

$$\begin{aligned}\tilde{H}(R_N, \theta, \lambda) = \frac{1-\phi(R_N, K)}{\phi_1'(R_N, K)}&[U'(W + R_N - \theta) - \lambda] + U(W + \tilde{M}(\lambda))\\ &- U(W + R_N - \theta) - \lambda[\tilde{M}(\lambda) + \theta + c - R_N].\end{aligned}$$

We have

$$\tilde{H}_1'(R, \theta, \lambda) = -\lambda\, r\,\frac{U(W + R) - U(W - B)}{U'(W + R)} < 0$$

and

$$\tilde{H}(\tilde{M}(\lambda) + \theta, \theta, \lambda) = -\lambda c < 0,$$

where r is the (constant) index of absolute risk aversion.

If $\tilde{H}(K, \theta, \lambda) > 0$ there exists $R_N^*(\theta, \lambda, K) \in (K, \tilde{M}(\lambda) + \theta)$ such that $\tilde{H}(R_N^*, \theta, \lambda) = 0$. Let $\tilde{R}_N$ be defined by

$$\begin{cases}\tilde{R}_N(\theta, \lambda, K) = R_N^*(\theta, \lambda, K) & \text{if } \tilde{H}(K, \theta, \lambda) > 0\\ \tilde{R}_N(\theta, \lambda, K) = K & \text{otherwise.}\end{cases}$$

$\tilde{R}_N$ maximizes $\mathcal{L}(R_N, \tilde{R}_A(\theta, \lambda, K), \lambda, \theta)$ w.r.t. R_N subject to $R_N \geq K$ when $\tilde{R}_A = \tilde{M}(\lambda) + \theta$. Furthermore, $\tilde{R}_N(\theta, \lambda, K)$ is continuous and almost everywhere continuously differentiable over the subset $\Theta \times R_+^* \times R_-^*$ where $\tilde{M}(\lambda) + \theta > K$. Finally $\tilde{R}_N(\theta, \lambda, K) \to K$ when $(\theta, \lambda, K) \to (\theta_0, \lambda_0, K_0)$ with $\tilde{M}(\lambda_0) + \theta_0 = K_0$.

Assume now that $\tilde{R}_A(\theta, \lambda, K) = K$. Let

$$\hat{H}(R_N, \theta, \lambda) = \frac{\mathcal{L}_1'(R_N, K, \lambda, \theta)}{\phi_1'(R_N, K) f(\theta)}.$$

We have $\hat{H}_1'(R_N, \theta, \lambda) < 0$ and

$$\hat{H}(K, \theta, \lambda) = \frac{1}{\phi_1'(K, K)}[U'(W + K - \theta) - \lambda] - \lambda c.$$

We have $U'(W + K - \theta) \leq \lambda$ when $\tilde{R}_A(\theta, \lambda) = K$, which gives $\hat{H}(K, \theta, \lambda) < 0$. Hence $\tilde{R}_N = K$ maximizes $\mathcal{L}(R_N, \tilde{R}_A(\theta, \lambda, K), \lambda, \theta)$ with respect to R_N subject to $R_N \geq K$ when $\tilde{R}_A = K$. $\tilde{R}_N(\theta, \lambda, K)$ is continuous and almost everywhere differentiable over $\Theta \times R_+^* \times R_-^*$.

Let $\tilde{C}(\lambda, K)$ be the expected cost associated with $\tilde{R}_N$ and $\tilde{R}_A$—that is, $\tilde{C}(\lambda, K)$ is obtained by substituting $\tilde{R}_N$ and $\tilde{R}_A$ in the left-hand side of (24). $\tilde{C}(\lambda, K)$ is continuous w.r.t. λ and K and decreasing w.r.t. λ. Furthermore $\tilde{C}(\lambda, K) \to +\infty$ when $\lambda \to 0$ and $\tilde{C}(\lambda, K) \to K$ when $\lambda \to +\infty$. Hence, when $K < 0$, there exists $\lambda(K) > 0$ such that $\tilde{C}(\lambda(K), K) = 0$ and $\lambda(K)$ is continuous over R_-^*. Hence $\{R_N(.), R_A(.)\}$ given by $R_N(\theta) \equiv \tilde{R}_N(\theta, \lambda(K), K)$ and $R_A(\theta) \equiv \tilde{R}_A(\theta, \lambda(K), K)$ is an optimal contract, conditionally on K when $K < 0$.

Step 2. Let $\Omega(K)$ be the corresponding expected utility of the policyholder. The results obtained at Step 1 show that $\Omega(.)$ is continuous over R_-^*.

When $K = 0$, the only feasible contract involves $R_N(\theta) = 0$ for all θ, which gives $\Omega(0) = EU(W - \theta)$. Let us show that $\Omega(K)$ is continuous at $K = 0$. Note first that $\Omega(K) \geq EU(W + K - \theta)$ since $R_N(\theta) \equiv K$ is feasible.

Let $\Omega^*(K)$ be the optimal expected utility when $c = 0$, with $\Omega(K) \leq \Omega^*(K)$. When $c = 0$, an optimal contract involves

$$\begin{aligned} R_A(\theta) &= R_N(\theta) = K \quad \text{if } \theta \leq \hat{\theta}(K) \\ R_A(\theta) &= R_N(\theta) = M + \theta \quad \text{if } \theta > \hat{\theta}(K), \end{aligned}$$

where $M = K - \hat{\theta}(K)$ and $\hat{\theta}(K)$ is continuous w.r.t. K. Hence, $\Omega^*(K)$ is continuous over R_-. Since

$$EU(W + K - \theta) \leq \Omega(K) \leq \Omega^*(K)$$

and

$$\Omega^*(0) = EU(W - \theta)$$

we deduce that

$$\Omega(0_-) = EU(W - \theta),$$

which implies that $\Omega(K)$ is continuous at $K = 0$.

Hence $\Omega(K)$ is continuous and reaches a maximum at $K^* \in [-B, 0]$. The contract $R_N(\theta) \equiv \tilde{R}_N(\theta, \lambda(K^*)K^*)$, $R_A(\theta) \equiv \tilde{R}_A(\theta, \lambda(K^*), K^*)$ is optimal.

Proof of (i). As shown in Step 1 above, $R_N(\theta) > K$ implies $R_A(\theta) = \tilde{M}(\lambda) + \theta > K$, hence the result with $M = \tilde{M}(\lambda)$. The fact that M is less than K will be established in the proof of (ii).

Proof of (ii). We know from the proof of Step 1 that at the optimum

$$\left.\begin{array}{r} R_N(\theta) = K \quad \text{if } \tilde{H}(K, \theta, \lambda) \le 0 \\ K < R_N(\theta) < M + \theta \quad \text{with} \quad \tilde{H}(R_N(\theta), \theta, \lambda) = 0 \quad \text{if } \tilde{H}(K, \theta, \lambda) > 0 \end{array}\right\} \tag{27}$$

Let us show that there exists $\theta^* > K - M$ such that $\tilde{H}(K, \theta^*, \lambda) = 0$. Assume that this is not the case. We have

$$\begin{aligned} \tilde{H}(K, K - M, \lambda) &= -\lambda c < 0 \\ \tilde{H}_1(R, \theta, \lambda) &< 0 \end{aligned}$$

and

$$\tilde{H}_2'(R, \theta, \lambda) = -\frac{[U(W + R) - U(W - B)]U''(W + R - \theta)}{U'(W + R)} + U'(W + R - \theta) - \lambda,$$

which implies, by using $M = \tilde{M}(\lambda)$, that $\tilde{H}_2'(R, \theta) > 0$ if $R < M + \theta$.

Hence $\tilde{H}(R, \theta, \lambda) < 0$ if $\theta \ge K - M$ and $R \ge K$, which gives $R_N(\theta) = K$ if $\theta \ge K - M$. We also have $R_N(\theta) = K$ if $\theta < K - M$. This means that the optimal contract provides no insurance (that is, $R_N(\theta) = K$ and $p(\theta) = 0$ for all θ is optimal) hence a contradiction. The characterization of $R_N(.)$ follows from (27) and $\tilde{H}_2' > 0$.

It remains to show that $\theta^* > 0$. Let $\omega(\theta)$ a Kuhn-Tucker multiplier associated with (25). The optimality conditions on $R_N(\theta)$ and K are respectively

$$\phi_1'(\theta)\tilde{H}(R_N(\theta), \theta, \lambda) + \omega(\theta) = 0 \quad \text{for all } \theta \tag{28}$$

$$\int_\Theta \phi_2(\theta)\{U(W + M) - U(W + R_N(\theta) - \theta) - \lambda[M + \theta + c - R_N(\theta)]\}\, dF(\theta) - \int_\Theta \omega_N(\theta)\, d\theta = 0 \tag{29}$$

with $M = \tilde{M}(\lambda)$. Note that, without loss of generality, we have assumed $R_A(\theta) = M + \theta$ for all θ. Indeed, when $R_N(\theta) = K$, we have $p(\theta) = \phi(K, K) = 0$ and $R_A(\theta)$ is arbitrary in such a case.

Integrating (28) multiplied by $\phi_2'(\theta)/\phi_1'(\theta)$ over Θ and substracting the result from (29) gives

$$\int_\Theta \frac{[1 - p(\theta)]\phi_2'(\theta)}{\phi_1'(\theta)}[U'(W + R_N(\theta) - \theta) - \lambda]\, dF(\theta) + \int_\Theta \frac{\omega(\theta)[\phi_2'(\theta) + \phi_1'(\theta)]}{\phi_1'(\theta)}\, d\theta = 0. \tag{30}$$

If $\omega(\theta) > 0$, then $R_N(\theta) = K$ and $\phi_2'(\theta) = -\phi_1'(\theta)$. Hence (30) and $M = \tilde{M}(\lambda)$ give

$$\int_\Theta \frac{[1 - p(\theta)]\phi_2'(\theta)}{\phi_1'(\theta)}[U'(W + R_N(\theta) - \theta) - U'(W + M)]\, dF(\theta) = 0, \tag{31}$$

which implies

$$\left.\begin{aligned} &\int_0^{\theta*} \frac{\phi_2'(\theta)}{\phi_1'(\theta)}[U'(W + K - \theta) - U'(W + M)]\, dF(\theta) \\ &\quad = \int_{\theta_+^*}^{\bar{\theta}} \frac{[1 - p(\theta)]\phi_2'(\theta)}{\phi_1'(\theta)}[U'(W + M) - U'(W + R_N(\theta) - \theta)]\, dF(\theta). \end{aligned}\right\} \tag{32}$$

The right-hand side of (32) is positive because of the concavity of $U(.)$. Assume that $\theta^* = 0$. Since there is a mass of probability at $\theta = 0$, (32) implies $K > M$, which contradicts the fact that $R_A(\theta) = M + \theta > K$ for all $\theta > \theta^*$. Hence we have $\theta^* > 0$. If $M \geq K$, we have $U'(W + K - \theta) - U'(W + M) > 0$ for all θ in $[0, \theta^*]$. We deduce that the left-hand side of (32) is negative, which is a contradiction. Hence, we have $M < K$.

Proof of (iii). Let $\eta(\theta) = M + \theta - R_N(\theta)$ for $\theta > \theta^*$ given by

$$\tilde{H}(M + \theta - \eta(\theta), \theta, \lambda) = 0$$

$\eta(\theta)$ is strictly positive. It is continuous since $H(R, \theta, \lambda)$ is almost everywhere continuously differentiable over $(K, +\infty) \times \Theta \times R_+$ and $\eta(\theta^*) = M + \theta^* - K$ since $H(K, \theta^*, \lambda) = 0$. We also have

$$R_N'(\theta) = 1 - \eta'(\theta) = -\frac{\tilde{H}_2'(R_N(\theta), \theta, \lambda)}{\tilde{H}_1'(R_N(\theta), \theta, \lambda)} > 0.$$

Let

$$\psi(R, K) \equiv \frac{1 - \phi(R, K)}{\phi_1'(R, K)} = \frac{U(W + R) - U(W - B)}{U'(W + R)}$$

with $\psi_1'(R, K) > 0$. We have

$$\begin{aligned}\tilde{H}(R_N(\theta), \theta, \lambda) = \; & \psi(R_N(\theta), K)[U'(W + M - \eta(\theta)) - \lambda] + U(W + M) \\ & - U(W + M - \eta(\theta)) - \lambda[c + \eta(\theta)] = 0 \quad \text{if } \theta > \theta^*,\end{aligned} \tag{33}$$

which implies

$$\begin{aligned}& [U'(W + M - \eta(\theta)) - \lambda]\psi_1'(R_N(\theta), K)R_N'(\theta) \\ & \quad = [\psi(R_N(\theta), K)U''(W + M - \eta(\theta)) - U'(W + M - \eta(\theta)) + \lambda]\eta'(\theta).\end{aligned} \tag{34}$$

The left-hand side of (34) is strictly positive when $\theta > \theta^*$, since $R_N'(\theta) > 0$, $\psi_1'(R_N(\theta), K)$ > 0 and $U'(W + M - \eta(\theta)) > U'(W + M) = \lambda$. The term into brackets in the right-hand side is strictly negative. Hence $\eta'(\theta) < 0$.

Finally, we have

$$U(W + M) - U(W + M - \eta(\theta)) - \lambda[c + \eta(\theta)] \in [A, B] \quad \text{for all } \theta > \theta^*,$$

with

$$\begin{aligned} A &= -\lambda[c + \eta(\theta^*)] \\ B &= U(W + M) - U(W + M - \eta(\theta^*)).\end{aligned}$$

When $U(.)$ is CARA, $\psi(R, K) \to +\infty$ when $R \to +\infty$. Hence (33) implies $U'(W + M - \eta(\theta)) \to \lambda = U'(W + M)$ when $\theta \to \infty$, which implies $\eta(\theta) \to 0$ when $\theta \to +\infty$. □

Proof of (iv). This results straightforwardly from Lemma 2 and from (ii) and (iii) in Proposition 2. □

Proof of Lemma 3

First, observe that there exists θ_0 given by (18) and (19) since $R_N(.)$ is supposed to be lower semicontinuous. Furthermore, Proposition 1 shows that the optimal contract satisfies all the constraints listed in Lemma 3. It remains to show that the contract obtained by maximizing $EU(W_f)$ subject to these constraints satisfies (3) for all $\theta, \hat{\theta}$ such that $\theta > \theta_0$ and $\hat{\theta} \neq \theta$.

Let $C(\theta, \hat{\theta})$ defined by

$$U(W + C(\theta, \hat{\theta}) - \theta) = [1 - p(\hat{\theta})]\, U\,(W - \theta + R_N(\hat{\theta})) + p(\hat{\theta})\, U\,(W - \theta - B) \quad \text{if } \theta \neq \hat{\theta}$$

and

$$U(W + C(\theta, \theta) - \theta) = [1 - p(\theta)]\, U\,(W - \theta + R_N(\theta)) + p(\theta)U(W - \theta + R_A(\theta)).$$

Equation (3) is equivalent to $C(\theta, \theta) \geq C(\theta, \hat{\theta},)$ if $\theta \neq \hat{\theta}$. If $U(.)$ is NIARA, we have $C(\theta, \hat{\theta}) \leq C(\theta', \hat{\theta})$ for all $\theta, \theta', \hat{\theta}$ such that $\theta > \theta', \hat{\theta} \neq \theta, \hat{\theta} \neq \theta'$.

Let θ and $\hat{\theta}$ such that $\theta > \theta_0$ and $\hat{\theta} \neq \theta$. The contract obtained by maximizing $EU(W_f)$ subject to the constraints listed in Lemma 3 is such that

$$C(\theta, \hat{\theta}) \leq C(\theta_0, \hat{\theta}) \leq R_N(\theta_0) \leq C(\theta, \theta).$$

The first above inequality results from $\theta > \theta_0$, the second one corresponds to (16) and the third one results from (15) and (18). We obtain $C(\theta, \theta) \geq C(\theta, \hat{\theta})$, which proves the desired result. □

Proof of Proposition 3

Proof of (i). Assume that for the optimal contract $\{R_A(.), R_N(.), p(.)\}$ there exist $\Delta_1 \subset \Theta$, $\Delta_2 \subset \Theta$ and a real number M such that for $i = 1, 2$: $\Delta_i \subset [\theta_0, \bar{\theta}]$, $\int_{\Delta_i} dF(\theta) > 0$ and $p(\theta) > 0$ if $\theta \in \Delta_i$.

Assume also that $R_A(\theta) < M + \theta$ if $\theta \in \Delta_1$ and $R_A(\theta) > M + \theta$ if $\theta \in \Delta_2$.

Consider another contract $\{\bar{R}_A(.), \bar{R}_N(.), \bar{p}(.)\}$ such that $\bar{p}(\theta) = p(\theta)$ and $\bar{R}_N(\theta) = R_N(\theta)$ for all θ, $\bar{R}_A(\theta) = R_A(\theta)$ if $\theta \notin \Delta_1 \cup \Delta_2$ and $\bar{R}_A(\theta) = R_A(\theta) + \varepsilon_i$ if $\theta \in \Delta_i$ for $i = 1, 2$, where ε_1 and ε_2 are small perturbations such that

$$\varepsilon_2 \int_{\Delta_2} p(\theta)\, dF(\theta) = -\varepsilon_1 \int_{\Delta_1} p(\theta)\, dF(\theta) > 0.$$

Lemma 3 implies that the new contract remains incentive compatible if ε_1 and ε_2 are small enough. Furthermore, the perturbations ε_1 and ε_2 do not modify the expected profit of the insurer. Finally, the first-order variation in the policyholder's expected utility is

$$\begin{aligned} dU &= \varepsilon_1 \int_{\Delta_1} p(\theta) U'(W - \theta + R_A(\theta))\, dF(\theta) \\ &\quad + \varepsilon_2 \int_{\Delta_2} p(\theta) U'(W - \theta + R_A(\theta))\, dF(\theta) \\ &\quad > U'(W + M) \left[\varepsilon_1 \int_{\Delta_1} p(\theta)\, dF(\theta) + \varepsilon_2 \int_{\Delta_2} p(\theta)\, dF(\theta) \right] = 0 \end{aligned}$$

Hence, the perturbation is feasible and welfare improving, which is a contradiction.

Proof of (ii). Assume $\theta_0 = 0$. Lemma 3 then implies that the optimal contract maximizes $EU(W_f)$ subject to Eqs. (2), (4), (6), (15), (16), and (18). This means that the optimal contract is an optimal solution to P_2 (under the additional constraint that $\tilde{\theta} = 0$, but this is unimportant since $\tilde{\theta} = 0$ at the optimal solution to P_2). We conclude that the characterization of the optimal contract given in Proposition 2 is still valid in this case.

Note that the proof of Proposition 2 does not use the fact that $U(.)$ is CARA except in the derivation $\tilde{H}_1'(R, \theta, \lambda)$. When $U(.)$ is NIARA, we have

$$\begin{aligned} \tilde{H}_1'(R, \theta, \lambda) = {} & [(r(W + R) - r(W + R - \theta))U'(W + R - \theta) - \lambda r(W + R)] \\ & \times \frac{U(W + R) - U(W - B)}{U'(W + R)}, \end{aligned}$$

where $r(W) \equiv -U''(W)/U'(W)$ is the index of absolute risk-aversion. When $U(.)$ is NIARA, we still have $\tilde{H}_1'(R, \theta, \lambda) < 0$ for all $\theta \geq 0$. Hence the proof of Proposition 2 remains valid in this case. □

Proof of Proposition 4

See Fagart and Picard [1998].

Acknowledgments

We are grateful to Bruno Jullien, Cuong Le Van, and the referees for useful comments on a preliminary version of this article.

Notes

1. Some of these questions are formulated by Townsend [1979]: "One might *conjecture*, based on the results for deterministic verification, that the probability of verification should be a nonincreasing function of y_2 and perhaps should be zero in states with high realizations." In Townsend's paper, y_2 may be interpreted as the wealth level of our policyholder: an increase in y_2 thus corresponds to a decrease in his level of loss. Likewise, Mookherjee and Png [1989] note that "the monotonicity of transfers and audit probabilities in the case without moral hazard, but with three or more possible income levels, remains an open question."
2. Mookherjee and Png [1989] assume that the consumption of the policyholder can be set to zero if the latter is detected to have lied. In an insurance context, assuming that feasible penalties are upward bounded is (hopefully) more realistic. Moreover, this assumption will make the problem much more tractable, particularly when the policyholder exhibits constant absolute risk aversion.
3. When B is the penalty (in monetary terms) incurred by a policyholder who is prosecuted after having been caught filing a fraudulent claim, then it may be more realistic to assume that the penalty is paid *in addition* to the premium P, since the latter is usually paid at the beginning of the time period during which the insurance contract is enforced. This assumption is equivalent to ours when the policyholder is affected by a liquidity constraint. Indeed, in such a case, an optimal contract specifies the largest possible premium compensated by a large insurance coverage, unless a fraudulent claim is detected by audit. This provides the best incentive to tell the truth and this does not affect equilibrium net payments.

   In this article, for the sake of simplicity, the penalty is supposed to be exogenously given, but optimal penalities could be characterized in a more general setting. In particular, imposing maximal penalties on defrauders may not be optimal either because auditing is imperfect (so that innocent policyholders may be wrongly indicted for loss misrepresentation) or because a policyholder may overestimate his damages in good faith. Furthermore, the literature on optimal auditing (particularly in the framework of tax compliance games) has emphasized the fact that very large penalties create incentives for policyholders to bribe auditors, which also limits the size of optimal penalties.
4. We assume that $R_N(.)$ is lower semicontinuous over Θ, which guarantees that Arg Min$\{R_N(\theta), \theta \in \Theta\} \neq \emptyset$.
5. Indeed, full insurance with $p(\theta) = 1$ for all θ dominates no-insurance if c is small enough. Hence the no-insurance contract is not optimal it c is small enough.
6. The reason is the following. If at the optimum $v(\theta) = U(W - \theta - B)$ for some level of loss θ, then the type-θ policyholder cannot be penalized in case of misreporting detected by audit. Hence, in such a case, the only way to deter fraudulent claiming is to verify with probability one all the claims that correspond to a net indemnity payment larger than $-B$. This means that the optimal audit policy is deterministic. In such a case, the insurance premium (which is paid in the no-audit region) would be equal to B. Assumption **A** is verified if the optimal net indemnity payment under deterministic auditing is larger than $-B$ for all θ. See Appendix A for details.

7. In (i) and (ii), a rigorous formulation is "for almost all θ" "—that is, for all θ except in a zero measure subset of Θ," instead of "for all θ." Likewise, "for some $\hat{\theta}$" in (iii) means "in a nonzero measure subset of Θ." We keep to this loose formulation for the sake of simplicity in the presentation. See Appendix B for details.
8. Note that $R_A(\theta)-$ respect. $R_N(\theta)-$ is arbitrary if $p(\theta) = 0-$ respect if $p(\theta) = 1-$. Hence, without loss of generality, we may assume that the optimal contract verifies $R_A(\theta) > R_N(\theta)$ for all θ.
9. Remind that $R_A(\theta) > R_N(\theta)$ at the optimal contract. Hence $dp(\theta) < 0$ implies $dR\theta > 0$.
10. See Townsend [1979], Bond and Crocker [1997], and Picard [1999].

References

BARON, D. and BESANKO, D. [1984]: "Regulation, Asymmetric Information and Auditing," *Rand Journal of Economics*, 15, 447–470.

BOND, E.W. and CROCKER, K.J. [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.

BORDER, K. and SOBEL, J. [1987]: "A Theory of Auditing and Plunder," *Review of Economic Studies*, 54, 525–540.

FAGART, M.-C. and PICARD, P. [1998]: "Optimal Insurance Under Random Auditing," working paper, THEMA 98–08.

GOLLIER, C. [1987]: "Pareto-Optimal Risk Sharing with Fixed Costs per Claim," *Scandinavian Actuarial Journal*, 62–73.

GUESNERIE, R. and LAFFONT, J.J. [1984]: "A Complete Solution to a Class of Principal-Agent Problems, with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.

HESTENES, M.R. [1980]: *Calculus of Variations and Optimal Control Theory*, Huntington, New York: Krieger Publishing.

HUBERMAN, G., MAYERS, D., and SMITH, JR., C.W. [1983]: "Optimum Insurance Policy Indemnity Schedules," *Bell Journal of Economics*, 14, 415–426.

MOOKHERJEE, D. and PNG, I. [1989]: "Optimal Auditing Insurance and Redistribution," *Quarterly Journal of Economics*, 104, 389–415.

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

TOWNSEND, R. [1979]: "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21, 265–293.