Insurance Law and Incomplete Contracts

Jean-Marc Bourgeon, Pierre Picard

Insurance Law and Incomplete Contracts

Jean-Marc Bourgeon* Pierre Picard†

July 5, 2018

Abstract

Under moral hazard, most insurance contracts are incomplete, to the extent that they condition the coverage neither on the contingencies under which policyholders choose their behavior, nor on the circumstances of the loss. This incompleteness can be explained by underwriting and auditing costs borne by insurers, by policyholders cognitive costs, and by the limits of market regulation. It opens the door to controversies and disputes between insured and insurer. In this context, we analyze how insurance law can mitigate moral hazard, by allowing insurers to cut indemnities in some circumstances, while preventing them from excessive nitpicking. We also highlight conditions under which the burden of proof should be on the policyholders, provided that insurers are threatened by bad faith penalties.

,

Keywords: insurance, moral hazard, incomplete contracts.
JEL Classification Numbers: D82, D86, G22.

*INRA and CREST- Ecole Polytechnique. Address: UMR Économie Publique, 16 rue Claude Bernard, 75231 Paris Cedex 05, France. E-mail: bourgeon@agroparistech.fr

†CREST- Ecole Polytechnique. Address: Department of Economics, Ecole Polytechnique, 91128 Palaiseau Cedex, France. E-mail: pierre.picard@polytechnique.edu. Pierre Picard acknowledges financial support from Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

1 Introduction

Rightly or wrongly, policyholders frequently think that insurers have leeway in settling claims, and that they nitpick on the indemnity payment if they believe it possible. This is particularly true for risk categories where insurance contracts and the soft-law guidelines provided by market regulators do not specify the coverage unambiguously in all possible contingencies. As a result, the insurer is perceived as having propensity for challenging the legitimacy of claims, by invoking provisions or practices of insurance law.

From a risk-sharing standpoint, this unpredictability of coverage weakens the efficiency of insurance contracting. On the other hand, conditioning coverage on the circumstances of the loss may be worthwhile under moral hazard if circumstances are informative concerning the policyholder's effort (Holmström, 1979; Shavell, 1979). However, in practice, the link between the circumstances of the claim and the indemnity is rarely specified in details in the insurance contract, and it is often limited to exclusions or force majeure clauses.[1] In other words, more often than not, insurance contracts are incomplete, and the resulting ambiguity may give a conflicted dimension to the insurer-insured relationship.[2]

This incompleteness can be attributed to the insurer's underwriting cost and to the policyholder's cognitive cost,[3] both of them being prohibitive in the case of complete

[1]For instance, a corporate property policy may exclude the damage resulting from fire caused by an explosion, or from the transportation of hazardous materials. Likewise, an insurer may impose on its customers to call on affiliated service providers (e.g., managed care networks), except force majeure. However, insurance contracts rarely include conditional indemnities depending on all possible circumstances, such as, for instance, the precise weather conditions at the origin of damages in the case of homeowner or automobile insurance, the number and quality (e.g., family or job-related connections) of witnesses, or the conditions when the policyholder is authorized to involve emergency services.

[2]The specificity of insurance contracts has provoked much attention from lawyers, judges and policymakers. This concerns primarily the incompleteness of contracts. As Abraham (1986) puts it: "Insurance policies often are not specific enough to make the rights and obligations of the parties during the claims process crystal clear". It has also been recognized that the sequential nature of the insurer-insured relationship (meaning that the insured has to pay the premium before coverage starts) puts policyholders in a position of inferiority, and may favor insurers' opportunism. In this regard, Works (1998) draws a parallel with the conditions identified by Williamson (1985), i.e., bounded rationality and asset specificity, under which opportunism is likely to thrive. He highlights how insurance law should limit the "forfeiture risk", i.e., the propensity of insurer to unduly deny coverage, by interpreting to their advantage the contractual conditions that trigger the payment of the insurance indemnity. This reinforces the importance of dispute resolution mechanisms in insurance law and regulation, including litigation, and, more often, arbitration, state-sponsored complaint-conciliation programs, and private ombudsmen schemes (see Schwarcz, 2008).

[3]For instance, Thaler & Sunstein (2009) and Handel & Kolstad (2015) present evidence that choosing a health insurance plan is viewed as a complicated decision by individuals, and that their choices are heavily influenced by factors such as context, switching costs, information frictions and

contracts. Furthermore, the operative event at the origin of the loss is frequently private information of the claimant, and exhaustively verifying the claim circumstances may be too costly. Observing the detailed circumstances of a loss requires a costly state verification process, and, most of the time, such an audit is justified only when the insurer has reason to believe that the policyholder misbehaved in some way.

The incompleteness of insurance contracts does not necessarily breaks the link between the indemnity paid to the claimant and the circumstances of the loss. In particular, the insurer may invoke legal means to deny coverage by elaborating on the evidence obtained by auditing the claim, which may be contested in court by the claimant. To take this forward, we will consider a model where insurers either validate claims on the basis of (freely available) soft information, or back denial of coverage by verifiable information obtained through a costly audit. Alleging some misconduct of the policyholder appears then as an indirect way to condition the insurance coverage on the circumstances of the loss, when the standard of proof used by courts is the balance of probabilities.[4]

Let us be more explicit about our approach. We will consider a setting where the insurer observes neither the action taken by the policyholder (as in usual moral hazard problems), nor the concrete contingencies she faced at that time, and that conditioned her behavior. This takes us away from a first-best world where, in the absence of transaction cost, risk averse policyholders should entirely transfer their risk exposure to risk-neutral insurers. The variety of states leads to a variety of behaviors among the population of policyholders. If a loss occurs, the insured files a claim that provides some partial information on the circumstances of the loss. In a second-best world with moral hazard, where the only asymmetry of information is about the policyholder's behavior, the optimal contract should condition the coverage on the state faced by the policyholder, and on the circumstances of the loss. However, although force majeure cases and some particular circumstances may be contractible, this second-best setting is not very realistic for a general approach to insurance contracting. In a third-best world, gathering evidence about the circumstances of the loss is costly and, furthermore, the insurer cannot know the contingencies that conditioned the policyholder's behavior. Hence, contracts are incomplete, i.e., its is impossible to spell out the relationship between the indemnity cut and the many states and loss circumstances. In this third-best setting, the insurer may decide to audit a claim in order to obtain evidence on

hassle costs.

[4]Demougin & Fluet (2006) show that this decision rule provides maximal incentives to exert care. See also Demougin & Fluet (2008) for an analysis of the case of imperfect evidence.

the circumstances of the loss, and possibly to allege that the policyholder misbehaved in some way. Such allegations may be contested by policyholders, and the law of insurance contracts will be the final arbiter of conflicts. When courts make decisions “on the balance of probabilities”, insurance law should allow insurers to cut indemnities in circumstances that are likely to reveal a severe misconduct of the policyholder. This leeway provided by law leads the insurer to indirectly condition coverage on the circumstances of the loss, which is a desirable feature of the contract but cannot be written in contractual clauses.

We will also investigate the burden of proof issue, when the costs of providing verifiable information (i.e., strong evidence that may checked by a court) are at the origin of an additional efficiency loss. We show that, in this fourth-best world, the burden of proof should be on the policyholder, if the cost of transmitting hard evidence is not larger (or, at least, not much larger) for the policyholder than for the insurer. Intuitively, attributing the burden of proof to the policyholder is a way to avoid the cost of transmitting hard evidence to the court, because the insurer has only to allege a misconduct that will not be contested at equilibrium. However, this is true only when the insurer faces a bath faith penalty if he is contradicted by the judge. In other words, giving the burden of the proof to the policyholder and threatening insurers with bad faith penalties appear to be the two arms of a balanced fourth-best solution.

Our analysis has many roots. The first one is the wide literature on incentives with incomplete contracts whose origin lies in the theory of the firm, when the state-contingent sharing of surplus cannot be exhaustively described in contractual arrangements between stakeholders (see Hart & Moore, 1999). It is also related to the analysis of conflicts in firms when labor contracts are incomplete because of the difficulty to gauge and verify the many aspects of performance such as teamwork or initiative.[5] Another approach to incomplete contracts focuses on the legal rules that restrict the set of feasible contracts and constrain the process of adversarial litigation in contract enforcement. It consists in analyzing incomplete contracts as agreements that do not specify what should be done by the parties in some contingencies and that include references to broad legal standards.[6] Our approach is linked to this second trend by considering a setting where insurers may refer to behavioral standards to deny claims, and by focusing attention on how insurance law and adversarial litigation restrict their

[5]See MacLeod (2007), or Malcomson (2012), for an overview of the “relational contracts” literature.

[6]See for instance Shavell (1980) on damage measures for breach of contract, and Scott & Triantis (2005) who challenges the stylized representation of legal enforcement in the concept of verifiability, and advocate a more sophisticated understanding of litigation in the analysis of contract design.

discretionary power. The analysis of the negligence rule by Fagart & Fluet (2009) is illustrative of this approach in the case of liability insurance.[7] A connection may also be made with the analysis of insurance fraud of Bourgeon & Picard (2014): by allowing insurers to cut the indemnity according to the circumstances of the loss, insurance law acts as an incentive device to perform costly audits.

The rest of the paper is as follows. Section 2 introduces our insurance model under moral hazard. Its main specificity, by comparison with more usual approaches, is twofold. Firstly, we consider an environment where the policyholder may be in various states when she chooses her behavior, hence a distribution of behaviors among a population of identical policyholders. Secondly, claims are characterized by the circumstances of the operative event (the accident) at the origin of the loss. In this setting, we characterize the optimal (second-best) insurance contract in a moral hazard setting, where the only asymmetry of information is about the policyholder's behavior. In Section 3, we turn towards an incomplete (third-best) contract setting, where verifying the circumstances of the loss requires a costly audit, while the state (i.e., the contingencies in which the policyholder has chosen her behavior) remains unknown to the insurer. We analyze the post-claim insurer-insured interaction in such a setting, with courts making decisions on the balance of probabilities. Section 4 focuses attention on a simplified case, with only two types of behavior ("effort" and "no effort"). This will allow us to highlight a trade-off between incentives through the rules of law and through partial insurance coverage that characterizes the third-best optimal solution. Section 5 considers the case where auditing claims only provides soft information about the circumstances of the loss, i.e., it just allows the insurer to take notice of these circumstances. Providing hard evidence requires verifiable information that can be transmitted to the court either by the policyholder or by the insurer, with specific additional costs. This leads us to a fourth-best optimal solution where the burden of proof may be given either to the insurer or to the insured. Section 6 concludes. Proofs are in an appendix.

[7] Fagart & Fluet (2009) consider a setting in which the level of care of a potential injurer is not verifiable, a signal about this behavior being perceived following the occurrence of harm affecting a victim. A non-contractible signal makes liability insurance contracts incomplete, but it nevertheless provides some evidence to courts. This evidence can be compared to a standard in order to implement the negligence rule, with a binary set of possible judicial decision {liable, non-liable}. Fagart & Fluet (2009) show that such an evidence-based negligence rule may Pareto dominate the strict liability rule.

2 The model

2.1 Notations and basic assumptions

Consider an insurance company providing coverage to a risk-averse individual (household or firm) against accidents that may result in a loss L. The occurrence of the loss depends on the policyholder's behavior which is indexed by $b \in \mathcal{B} = \{1, 2, \ldots, n\}$ and ranks the probability of accident π_b increasingly, i.e. $\pi_1 < \pi_2 < ... < \pi_n$. Hence, $b = 1$ corresponds to a cautious behavior with the lowest probability of accident, and the other behaviors $b \in \{2, ..., n\}$ correspond to various types of misconducts, increasingly risky, but also decreasingly demanding in terms of effort. The disutility of each behavior b is however imperfectly known to the policyholder at the time she takes out the insurance policy: it depends on a parameter θ that reflects the diversity of concrete situations in which she may find herself during the policy period, and which we refer to as the "state" in which the individual finds herself when she chooses her behavior.[8] We assume $\theta \in \Theta$, where Θ is the (multidimensional) set of possible states, and the disutility of behavior b in state θ is denoted $d_b(\theta)$, with $d_b(\theta) > d_{b+1}(\theta)$, for all $b \in \mathcal{B}$ and all $\theta \in \Theta$. Hence, in all states, less risky behaviors entail a larger disutility because they require more effort on the part of the policyholder. We assume that θ is distributed in Θ according to a continuous c.d.f. $H(\theta)$ with density function $h(\theta)$. If the policyholder chooses behavior b in state θ, then her utility is $u(W_f) - d_b(\theta)$ where W_f is her final wealth and u is a (twice continuously differentiable) von Neumann-Morgenstern utility function such that $u' > 0, u'' < 0$.

The insurer may collect evidence on the circumstances of the accident, which leads to an index $x \in [0, 1]$ that reflects the more or less risky behavior of the policyholder, with $x = 0$ when there is no claim. If a claim is filed, x is continuously distributed in $(0, 1]$ and measures the likelihood of the various misconducts that may have been at the origin of the claim. The distribution of x depends on the policyholder's behavior $b \in \mathcal{B}$. It is distributed over $[0, 1]$ according to c.d.f. $G_b(x)$, with a mass of probability $G_b(0) = 1 - \pi_b$ corresponding to the no-accident case, and density $g_b(x) = G'_b(x)$ if

[8] For instance, a car driver may exert a low level of effort because she does not adequately maintain her vehicle, or because her speed is not appropriate, or because she drives after drinking, or because of a mixture of these behaviors, and all possible misconducts correspond to $b \in \{2, ..., n\}$. θ may be viewed as a random shock that affects the driver's disutility of refraining from these various misconducts. For example, vehicle maintenance will be more painful to the owner when she is struggling to make ends meet in unexpected tough economic circumstances. Likewise, a driver may think that it might not be so bad to break the speed limit if it is a question of arriving on time at an important business meeting.

$x \in (0,1]$. We denote $\hat{g}_b(x) = g_b(x)/\pi_b$ the density of x for behavior b, conditionally on a loss occurring. We assume that an increase in b induces a shift in the distribution of x, with strict MLRP when a loss occurs, i.e. $g_{b+1}(x)/g_b(x)$ is increasing with respect to x in $(0,1]$ for all $b \in \{1,\ldots,n-1\}$. In words, a larger x leads one to think that the policyholder was at fault of a more severe misconduct.[9,10] By an abuse of language, in what follows we may refer to x as the circumstances of the loss, although this is only the suspicion index derived from them. Hence, state θ refers to everything that conditions the policyholder's decision making about her behavior b, while circumstances x characterize the operative event at the origin of the financial loss L.

An insurance contract specifies a premium P paid at the outset, and a non-negative indemnity I in the case of a loss. We denote by $b^\star(\cdot) : \Theta \to \mathcal{B}$ the corresponding policyholder's behavioral rule, i.e. $b^\star(\theta)$ is the action taken by the insured in state $\theta \in \Theta$.[11,12] We neglect any transaction cost, and assume that insurers are risk neutral. Hence, the insurer's break-even constraint imposes that the insurance premium covers the expected indemnity payments. We also assume that over-insurance is ruled out, either for legal reasons, or because the policyholder could deliberately create losses in order to pocket the insurance indemnity. An optimal insurance contract maximizes the individual's expected utility (i.e., her expected utility before she knows her state θ) under the insurer's break-even constraint, the no-overinsurance constraint and the non-negative indemnity constraint.

In such a setting, a first-best allocation corresponds to the case where there is no asymmetry of information of any kind between insurer and insured. In particular, the insurer observes the state θ and the policyholder's behavior b, and this information is verifiable by a third party, like a court. It is well-known that such a first-best allocation

[9]We may be more explicit, and denote by $\omega \in \Omega$, the operative event at the origin of the loss, where Ω is the set of all possible events (i.e., all types of accidents that may occur). Let $(\Omega, \mathcal{F}, \mathbb{P}_b)$ be a probability space, with a probability measure $\mathbb{P}_b$ for each policyholder's behavior $b \in \mathcal{B}$. We know from Milgrom (1981) that we can associate a real variable x to event ω through a function $x = \varphi(\omega)$, with $\varphi : \Omega \longrightarrow [0,1]$, such that x is a sufficient statistic for b and satisfies MLRP for all b. In particular, for any nondegenerate prior on b, an increase in x induces a FOSD shifts in the posterior probability distribution of b. In this sense, a larger x can be interpreted as a "bad news", i.e., as suggesting that the policyholders is guilty of a more serious misconduct.

[10]To simplify matters (and to avoid corner solutions), we also assume $g_n(x)/g_1(x) \to 0$ when $x \to 0$ and $g_n(x)/g_1(x) \to \infty$ when $x \to 1$. Intuitively, $b = 1$ is much more likely than $b = n$ when x is close to 0, and conversely when x is close to 1.

[11]In particular, an insurance policy inducing a behavior $b^\star(\theta) = 1$ for all θ is usually suboptimal because $d_1(\theta)$ may be very large in some states θ. To take an extreme example, think of a man who breaks the speed limit when driving his wife to the hospital maternity, or, less dramatically, think of a driver who is worrying about arriving late at a business meeting.

[12]In what follows, we consider behavioral rules for which incentives matters, in the sense that $b^*(\theta) < n$ in a positive-measure subset of Θ.

is characterized by full coverage $I = L$ whatever the state θ and the circumstances of the loss x.

2.2 Second-best contract under moral hazard

A second-best solution to the optimal insurance contracting problem corresponds to the case where the insurer does not observe the policyholder's behavior b, which is thus private information to the policyholder, but there is no other asymmetry of information between insured and insurer. In particular, both of them observe the state θ, and the circumstances x of any accident which may occur. This information (θ, x) is costlessly verifiable and hence it can condition the insurance coverage. Let $I(\theta, x)$ denotes the insurance indemnity. The policyholder's final wealth is $W_f = W - P$ if there is no accident, and $W_f = W - P - L + I(\theta, x)$ when an accident with circumstances x occurs in state θ. Thus, conditionally on state $\theta \in \Theta$, the expected utility of a policyholder with behavior $b \in \mathcal{B}$ is written as $u_b(\theta) - d_b(\theta)$, where

$$u_b(\theta) = (1 - \pi_b)u(W - P) + \pi_b \int_0^1 u(W - P - L + I(\theta, x))\hat{g}_b(x)dx. \tag{1}$$

The first and second terms in (1) correspond to the no-accident and accident states respectively. If the insurance contract induces behaviors $b^\star(\theta)$ in state θ, then the ex ante expected utility of the policyholder (when she signs the contract) and her ex post incentive constraints (once she has learned about the relevant θ) are written as

$$\mathbb{E}u^\star \equiv \int_\Theta [u_{b^\star(\theta)}(\theta) - d_{b^\star(\theta)}(\theta)]dH(\theta), \tag{2}$$

and

$$u_{b^\star(\theta)}(\theta) - d_{b^\star(\theta)}(\theta) \geq u_b(\theta) - d_b(\theta) \text{ for all } (\theta, b) \in \Theta \times \mathcal{B}, \tag{3}$$

respectively. We neglect any transaction costs, and the insurance premium must at least cover the expected indemnity payments, i.e.

$$P \geq \int_\Theta \pi_{b^\star(\theta)} \int_0^1 I(\theta, x)g_{b^\star(\theta)}(x)dxdH(\theta). \tag{4}$$

Finally, the no-overinsurance and the non-negativity constraints are written as

$$0 \leq I(\theta, x) \leq L \text{ for all } (\theta, x) \in \Theta \times [0, 1]. \tag{5}$$

Under a competitive insurance market, the policyholder obtains the whole surplus of her relationship with the insurer. The optimal insurance contract maximizes the expected utility given by (2) with respect to P, $I(\cdot)$ and $b^{\star}(\cdot)$ subject to (3),(4) and (5).

Unsurprisingly, as shown in Proposition 1, $I^{\star}(\theta, x)$ actually depends on θ and x. Indeed, for a given behavioral rule $b^{\star}(\cdot)$, the uncertainty about θ should be taken into account in the insurance coverage.[13] Furthermore, x is informative about the policyholder's effort in the sense of Holmström (1979) and, in our moral hazard context, it should condition the transfer from insurer to policyholder. Proposition 1 shows how this conditioning should be implemented.

Proposition 1 *For all $\theta \in \Theta$ such that $b^{\star}(\theta) < n$, there exist $\underline{x}(\theta), \overline{x}(\theta) \in [0,1]$ with $\underline{x}(\theta) < \overline{x}(\theta)$, such that the second-best optimal indemnity schedule $I^{\star}(\theta, x)$ is continuous in x, with*

$$\begin{array}{ll} I^{\star}(\theta,x) = L & \text{if } 0 \leq x < \underline{x}(\theta) \text{ if } \underline{x}(\theta) > 0, \\ \left.\begin{array}{l} 0 < I^{\star}(\theta,x) < L \\ dI^{\star}(\theta,x)/dx < 0 \end{array}\right\} & \text{if } \underline{x}(\theta) < x < \overline{x}(\theta), \\ I^{\star}(\theta,x) = 0 & \text{if } \overline{x}(\theta) < x \leq 1 \text{ if } \overline{x}(\theta) < 1. \end{array}$$

If $b^{\star}(\theta) = n$, then $dI^{\star}(\theta, x)/dx = 0$ for all x.

In all states θ where some effort is required (i.e., $b^{\star}(\theta) < n$), the optimal insurance policy provides full coverage, partial coverage or zero coverage, depending on the circumstances of the loss. The more favorable the circumstances (i.e., the lower x), the larger the indemnity. The higher and lower bound L and 0 may be reached under the most favorable or worst possible circumstances (i.e., when $0 \leq x \leq \underline{x}(\theta)$ and $\overline{x}(\theta) \leq x \leq 1$), respectively. There is partial coverage in the intermediary cases, with larger coverage when circumstances are more favorable. Equivalently, we may write $I^{\star}(\theta, x) = [1 - z^{\star}(\theta, x)]L$, where $z^{\star}(\theta, x)$ is an indemnity cut such that $z^{\star}(\theta, x) = 0$ if $x < \underline{x}(\theta)$, $z^{\star}(\theta, x) \in (0, 1)$ with $z^{\star}(\theta, x) > 0$ if $\underline{x}(\theta) < x < \overline{x}(\theta)$ and $z^{\star}(\theta, x) = 1$ if $x > \overline{x}(\theta)$. These results are illustrated in Figure 1.

[Figure 1 about here.]

[13] By way of follow-up to footnote 11, it is conceivable that, when a road traffic offense has been committed, cutting the indemnity is less recommended in the case of the man who drives his wife to the maternity in emergency, than when the same person is worried about arriving late at a business meeting. The behavioral disutility associated with obeying speed limits is probably larger in the first case than in the second.

By comparison, we may consider the two polar cases where circumstances x are either totally uninformative or perfectly informative about behavior b. Circumstances do not convey any information on the policyholder's conduct when $\hat{g}_b(x) = 1$ for all $x \in (0, 1]$ and all $b \in \mathcal{B}$. In this case, the optimal insurance indemnity does not depend on x, and partial coverage is optimal when the policyholder should be incentivized to exert effort. In other words, we should have $I^\star(\theta, x) = \overline{I}(\theta) \in (0, L]$, with $\overline{I}(\theta) < L$ if $b^\star(\theta) < n$.[14] This corresponds to the most common setting of insurance contracts under moral hazard. Conversely, signal x is perfectly informative when the policyholder's behavior can be deduced from the circumstances of the loss. This is the case if there is a partition of interval $[0, 1]$, each subinterval being associated with a particular behavior b. More explicitly, there exists a sequence $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{n+1}$ such that $\tilde{x}_1 = 0, \tilde{x}_{n+1} = 1$ and $\tilde{x}_b < \tilde{x}_{b+1}$ for all $b \in \mathcal{B}$, such that $g_b(x) > 0$ if $x \in [\tilde{x}_b, \tilde{x}_{b+1})$ and $g_b(x) = 0$ otherwise. In that case, in state θ, the policyholder is incentivized to choose the first-best behavioral rule $b^\star(\cdot)$ by a stepwise decreasing indemnity schedule with full coverage $I^\star(\theta, x) = L$ if $x \in [0, \tilde{x}_{b^\star(\theta)+1})$ and zero coverage $I^\star(\theta, x) = 0$ if $x \in [\tilde{x}_{b^\star(\theta)+1}, 1]$. If the circumstances of the loss are partially informative about the policyholder's conduct, then the optimal second-best indemnity schedule reaches a compromise between these two polar cases: it is continuously decreasing, possibly with full coverage under the most favorable circumstances, and no coverage under highly unfavorable ones.

3 Incomplete contracts

Except in exceptional cases (e.g., a well-documented natural disaster being at the origin of force majeure), the state θ in which policyholders find themselves is not observed by insurers, and verifying it by audit would be very difficult, and even impossible in many cases. As regards the circumstances of the loss x, there are at least two reasons for which, in practice, they may condition the insurance coverage only in a very rudimentary way, particularly though exclusions for specific types of accidents.[15] Firstly, obtaining verifiable information about the circumstances of an accident usually requires a costly verification process. More often than not, claimhandlers routinely pay the insurance indemnity, but sometimes a (privately perceived) signal convinces the insurer that the claim should be audited in order to know more about the circumstances of the claim. Equal treatment of policyholders prevents insurers conditioning coverage

[14] In cases of force majeure, it may be optimal not to exert effort (i.e., choosing $b = n$), but these are truly exceptional situations.

[15] For instance, a corporate fire insurance policy may exclude damages caused by explosion.

on the information obtained by audit for some of them. Another reason is related with the cognitive costs of contracts. It is true that insurance policies frequently contain a lot of small print, but, apart from exclusion clauses, they mainly focus either on the law principles governing insurance contracts, or on policyholders' duties.[16] In fact, property and liability losses may occur through operative events that correspond to such a large number of circumstances that listing and describing them all, with specific coverage in all cases, would be unfeasible in practice.[17]

Hence, in most cases, contractual insurance payments are neither conditioned upon the ex ante situation of the policyholders nor on the ex post circumstances of the loss, although such circumstances may nevertheless be verified by audit. It remains no less true that the insurer may use the evidence yielded by audits to invoke a breach of law in order to justify a more or less severe cut in the indemnity. Hence, auditing the claim may be the starting point of disputes between insurer and insured that are resolved through some legal arrangement, be it an amicable settlement, by resorting to an arbitrator or by going to court. This is the interaction between the verification of the circumstances of the loss and the stipulations of insurance law that we will contemplate in what follows. The auditing cost coupled with the incompleteness of insurance policies are obviously at the origin of an efficiency loss, by comparison with the second-best solution. We thus consider here a third-best solution to the insurance problem under moral hazard.

Let us follow a standard way in the analysis of conflicts arbitrated by law, which consists of assuming that judges decide by relying on the likelihood of the behavior alleged by each party. An insurer may allege that the policyholder misbehaved and thus, on the basis of the law of contracts, that the claim should be fully or partially denied. However, the insurer's allegations must be consistent with the empirical evidence provided by the circumstances of losses, for otherwise the judge would consider them as insurer's bad faith and they would be invalidated. In other words, insurance law may allow insurers to condition insurance payments on loss circumstances by opening the door to legal indemnity cuts ("law completes contracts" as it is sometimes said), but its application is constrained by the approval of the judge on the basis of available information. This corresponds to the usual standard of proof for civil cases in the Common Law: judges are supposed to decide "on the balance of probabilities".

[16] For instance, calling for the service of affiliated car repairers or health care providers.

[17] Put differently, by using the terminology of cognitive science, such complete contracts would induce such a large cognitive load (i.e., too much effort in using working memory) that the incentive advantages of such contracts are not worth it.

Assume that, when a claim is filed, the insurer privately observes a signal $s \in \mathbb{R}$ defined by

$$s = x + \varepsilon,$$

where ε is a zero-mean random variable, with $Cov(x, \varepsilon) = 0$. We assume that x can be verified by auditing the claim, which costs c to the insurer, and we denote $q(s) \in [0, 1]$ the audit probability when signal s is perceived.

When no audit is performed, the insurer routinely pays I to the claimant, and in that case we have $W_f = W - P - L + I$. If x has been verified through an audit, then the insurance payment depends upon x through legal means that may be invoked by the insurer. For notational simplicity we will not distinguish misconduct $b \in \{2, \ldots, n\}$ from the corresponding legal means (or broad standards) that can be invoked by insurers, although, in practice, there are many types of misconduct, while the law of insurance contracts only includes a limited number of legal means.[18] Insurance law specifies the insurer's leeway in the claim settlement process, i.e. to what extent a legal means allows him to cut or even cancel coverage. More precisely, in what follows, the law of insurance contracts is subsumed in the proportions of claims $y_b \in [0, 1]$ that the insurer is allowed to cut for each behavior b. When the insurer is allowed to fully cancel the claim, we have $y_b = 1$. We assume $y_1 = 0$ because the insurer can cut the indemnity only by alleging that the policyholder misbehaved in some way, and that the law is constrained by a severity principle, according to which the severity of misconducts and the intensity of indemnity cuts are co-monotone, i.e., $y_b \leq y_{b+1}$ for all $b = 1, \ldots, n-1$. If the insurer is in a position to invoke legal means b under circumstances x, then he may decide to cut the indemnity by a fraction $z(x)$ lower or equal to y_b, and the insurance payment is $[1 - z(x)]I$, with $W_f = W - P - L + [1 - z(x)]I$

When the policyholder behaves according to the state-dependent rule $b^\star(\theta)$, Bayes Law provides the conditional probability of a behavior b when signal $x \in (0, 1]$ is observed after an audit. This is written as

$$\Pr(b|x) = \frac{g_b(x) \int_{\Theta_b^\star} dH(\theta)}{\sum_{b' \in \mathcal{B}} g_{b'}(x) \int_{\Theta_{b'}^\star} dH(\theta)}, \tag{6}$$

where $\Theta_b^\star \equiv \{\theta \in \Theta | b^\star(\theta) = b\}$ is the set of states in which the policyholder chooses $b \in \mathcal{B}$.

Given x, alleging misconduct $b_0 \in \{2, \ldots, n\}$ is said to be "credible on a balance

[18] The list of available legal means include: non-disclosure, misrepresentation, duty of care, reckless conduct, bad faith... with possible overlaps.

of probabilities" if it is more likely that the policyholder had misconduct b_0 or a worse misconduct $b \in \{b_0+1,\ldots,n\}$ than a better behavior $b \in \{1,\ldots,b_0-1\}$, i.e., if[19]

$$\sum\nolimits_{b=b_0}^{n} \Pr(b|x) \geq \sum\nolimits_{b=1}^{b_0-1} \Pr(b|x),$$

or, equivalently if

$$\sum\nolimits_{b=b_0}^{n} \Pr(b|x) \geq \frac{1}{2}.$$

We denote $\hat{b}(x)$ the most serious misconduct that can be credibly alleged when signal x is perceived through an audit, i.e.,

$$\sum\nolimits_{b=\hat{b}(x)+1}^{n} \Pr(b|x) < \frac{1}{2} \leq \sum\nolimits_{b=\hat{b}(x)}^{n} \Pr(b|x), \tag{7}$$

with $\hat{b}(x) = 1$ if no misconduct $b_0 \in \{2,\ldots,n\}$ is credible. The following lemma shows that a larger x allows the insurer to credibly allege more serious misconducts.

Lemma 1 $\hat{b}(\cdot) : [0,1] \to \mathcal{B}$ *is a non-decreasing step function.*

Lemma 1 allows us to write $\hat{b}(x) = b$ if $x_b \leq x < x_{b+1}$, for all $b \in \mathcal{B}$, with $x_1 = 0$ and x_b given by

$$\sum\nolimits_{b'=b}^{n} \Pr(b'|x_b) = \frac{1}{2}$$

for all $b = 2, ..., m$, where $m \in \mathcal{B}$ is the most severe misconduct that can be confirmed by the judge, i.e., that is credible under the balance of probabilities when x is close to 1. The corresponding maximum indemnity cuts are given by $y_{\hat{b}(x)}$ as illustrated Figure 2 (where $m = 4$).

[Figure 2 about here.]

The policyholder-insurer interaction can then be described by the following five-stage game:

- **Stage 1:** The individual takes out the insurance policy (I, P). Nature chooses θ. The policyholder observes θ and then she chooses behavior $b \in \mathcal{B}$. Should a loss occur, she files a claim. In that case, the insurer observes signal s.

[19] Observe that this legal standard, also called "preponderance of evidence", is generally understood as implying a threshold degree of certainty just above 50%. Other legal standards, like "proof beyond a reasonable doubt" or "clear and convincing evidence," would correspond to different (larger) levels of certainty that could be easily dealt with in our setup, without changing our results qualitatively.

- **Stage 2:** The insurer either directly validates the claim or triggers an audit. In that case, he incurs the audit cost c and he gets the verifiable information x.

- **Stage 3:** If an audit has been performed, the insurer either validates the claim and pays I to the claimant, or he alleges that the policyholder misbehaved according to $b \in \{2, \ldots, n\}$.

- **Stage 4:** The policyholder may decide to contest in court the insurer's allegation. The judge confirms the insurer's allegation if $b \leq \hat{b}(x)$, and he dismisses it otherwise.

- **Stage 5:** The indemnity paid to the policyholder is I if the claim has been validated by the insurer or if the insurer's allegation b has been dismissed by the judge. Otherwise, the insurer pays an indemnity $(1-z)I$, with $z \leq y_b$.

A subgame perfect equilibrium of this game is easily characterized. After observing s at stage 1, the insurer triggers an audit at stage 2 if $c \leq \mathbb{E}[y_{\hat{b}(x)}|s]I$, since $b = \hat{b}(x)$ is the most severe allegation made at stage 3 that will not be dismissed by the judge at stage 4, and the insurer chooses $z = y_{\hat{b}(x)}$ at stage 5. Thus, an equilibrium audit strategy is defined by

$$q(s) = \begin{cases} 1 & \text{if } c \leq \mathbb{E}[y_{\hat{b}(x)}|s]I, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Lemma 2 *The equilibrium audit strategy is a unit step function: $q(s) = 0$ if $s < s^\star$ and $q(s) = 1$ if $s \geq s^\star$, with $s^\star \in \mathbb{R} \cup \{-\infty, +\infty\}$.*

Intuition of lemma 2 is straightforward: claims should be audited when the signal s is bad enough to be considered as a red flag, indicating that the circumstances of the loss are likely to be unfavorable (i.e., x is probably large). Cases where $s^\star = \pm\infty$ correspond to corner solutions where claims are never (resp. always) audited because c is very large (resp. very low).

Conditionally on state $\theta \in \Theta$, the expected utility of a policyholder with behavior $b \in \mathcal{B}$ is written as $u_b(\theta) - d_b(\theta)$, where

$$\begin{aligned} u_b(\theta) = {} & (1-\pi_b)u(W-P) + \pi_b \int_0^1 u(W-P-L+I)\left(1 - \mathbb{E}[q(s)|x]\right)\hat{g}_b(x)dx \\ & + \pi_b \int_0^1 u(W-P-L+(1-z(x))I)\mathbb{E}[q(s)|x]\hat{g}_b(x)dx. \end{aligned} \tag{9}$$

The first, second and third terms in (9) correspond to the no-accident state, to the accident states without audit and to the accident states with audit, respectively.

For an optimal insurance law, the insurance contract (P, I), the audit strategy $q(\cdot)$, and the insurance law $\{y_b, b \in \mathcal{B}\}$ maximize

$$\mathbb{E}u^\star \equiv \int_\Theta [u_{b^\star(\theta)}(\theta) - d_{b^\star(\theta)}(\theta)]dH(\theta) \tag{10}$$

subject to

$$P \geq \int_\Theta \pi_{b^\star(\theta)} \left\{ I + \int_0^1 (c - Iy_{\hat{b}(x)})\mathbb{E}[q(s)|x]\hat{g}_{b^\star(\theta)}(x)dx \right\} dH(\theta), \tag{11}$$

$$u_{b^\star(\theta)}(\theta) - d_{b^\star(\theta)}(\theta) \geq u_b(\theta) - d_b(\theta) \text{ for all } (\theta, b) \in \Theta \times \mathcal{B}, \tag{12}$$

$$q(s) = \begin{cases} 1 & \text{if } s \geq s^\star \\ 0 & \text{if } s < s^\star \end{cases}, \tag{13}$$

where $s^\star$ is given by

$$c = \mathbb{E}[y_{\hat{b}(x)}|s^\star]I, \tag{14}$$

and $\hat{b}(\cdot) : [0,1] \to \mathcal{B}$, $b^\star(\cdot) : \Theta \to \mathcal{B}$ satisfy conditions (6) and (7).

Notations can be recapped as follows: $b^\star(\theta)$ is the policyholder's behavior in state θ, $\hat{b}(x)$ is the most severe policyholder's misconduct that can be alleged by the insurer after observing circumstances x through a claim audit, and $s^\star$ is the signal threshold above which an audit is triggered. Hence, $\mathbb{E}u^\star$ given by (10) is the ex ante policyholder's expected utility, i.e., before she knows the state θ, with behavioral rule $b^\star(\cdot)$. Condition (13) is deduced from condition (8) and Lemma 2: an audit is triggered if $s \geq s^\star$, where $s^\star$ is given by (14). Auditing a claim costs c, but it allows the insurer to reduce the indemnity from I to $I[1 - y_{\hat{b}(x)}]$ if the audit reveals claim circumstances x. Hence, (11) is the insurer's break-even condition. (12) is the ex post incentive constraint: in state θ, the policyholder weakly prefers behavior $b^\star(\theta)$, rather than any other behavior $b \neq b^\star(\theta)$.

Proposition 2 *If c is not prohibitively large, then an optimal third-best solution to the insurance moral hazard problem is such that $y_b > 0$ in a non-empty subset of $\mathcal{B}$ and the insurer audits claims with positive probability (i.e., $s^\star < +\infty$).*

In short, the insurer can allege that the policyholder misbehaved in some way (i.e., $b \geq 2$) when x is sufficiently large without being dismissed by the court, and in such cases, insurance law should allow him to cut indemnities. This leeway provided by the law leads the insurer to condition the payment on the circumstances of the loss, which is a desirable feature of the relationship between insurer and insured, but cannot be written in contractual clauses. This will be sustained by an equilibrium strategy where claims are verified with positive probability (i.e., when s is large enough) if the audit cost c is not too large.

4 The two-type case

In this section, we characterize more precisely the optimal second-best insurance contract in the simple case where $n = 2$, with a single type of misconduct ($b = 2$), and two states, i.e. $\Theta = \{\underline{\theta}, \bar{\theta}\}$ with $d_1(\underline{\theta}) > d_2(\underline{\theta}) > 0$ and $d_1(\bar{\theta}) = +\infty, d_2(\bar{\theta}) > 0$. In words, $\underline{\theta}$ is a normal state in which the policyholder can be incentivized in order to choose $b = 1$, while she never chooses $b = 1$ when she is in state $\bar{\theta}$. We consider an optimal second-best allocation where the policyholder chooses $b = 1$ when $\theta = \underline{\theta}$, and $b = 2$ when $\theta = \bar{\theta}$. We denote $\underline{h} = \Pr\{\theta = \underline{\theta}\}$ and $\bar{h} = \Pr\{\theta = \bar{\theta}\}$ the probabilities of the states, with $\underline{h} + \bar{h} = 1$. The law of contracts allows the insurer to reduce the indemnity by a proportion y_2 when $x \geq x_2$ where the threshold x_2 is deduced from $\Pr\{b = 2|x_2\} = 1/2$. We denote $\hat{S}_b = \Pr\{x \geq x_2|b\} = 1 - \hat{G}_b(x_2)$, where $\hat{G}_b$ is the c.d.f. of x, conditionnally on a loss occuring under behavior b. The optimal solution of our problem is easier to characterize by defining

$$\begin{aligned} \hat{u} &= u(W - P), \\ u_0 &= u(W - P - L + I), \\ u_1 &= u(W - P - L + (1 - y_2)I), \end{aligned}$$

and by denoting

$$\begin{aligned} v &= u_0 - u_1, \\ \bar{g} &= 1 - [\pi_2 \hat{G}_2(x_2) - \pi_1 \hat{G}_1(x_2)]/(\pi_2 - \pi_1), \\ A &= [d_1(\underline{\theta}) - d_2(\underline{\theta})]/(\pi_2 - \pi_1) > 0. \end{aligned}$$

Since the type-$\underline{\theta}$ incentive constraint is obviously binding, we may rewrite this constraint as

$$\hat{u} = u_0 + A - v\bar{g}. \tag{15}$$

Substituting

$$P = W - u^{-1}(\hat{u}) = W - u^{-1}(u_0 + A - v\bar{g}), \tag{16}$$

$$I = P + L - W + u^{-1}(u_0), \tag{17}$$

$$(1 - y_2)I = P + L - W + u^{-1}(u_0 - v), \tag{18}$$

in the insurer's break-even constraint gives

$$\begin{aligned} \bar{W} &\geq (\underline{h}\pi_1\hat{G}_1 + \bar{h}\pi_2\hat{G}_2)u^{-1}(u_0) + (\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)u^{-1}(u_0 - v) \\ &+ (1 - \underline{h}\pi_1 - \bar{h}\pi_2)u^{-1}(u_0 + A - v\bar{g}) \\ &\equiv K(u_0, v), \end{aligned} \tag{19}$$

where $\bar{W} = W - L(\underline{h}\pi_1 + \bar{h}\pi_2)$ is the ex ante policyholder's expected wealth. Since u is concave, u^{-1} is convex and thus $K(u_0, v)$ is a convex function of u_0 and v, with $\partial K/\partial u_0 > 0$ and $\partial K/\partial v < 0$. The same substitutions yield

$$\mathbb{E}u^\star = \underline{h}[(1 - \pi_1)\hat{u} + \pi_1(u_0 - \hat{S}_1 v)] + \bar{h}[(1 - \pi_2)\hat{u} + \pi_2(u_0 - \hat{S}_2 v)] - \underline{h}d_1(\underline{\theta}) + \bar{h}d_2(\bar{\theta})$$

and, using (15) and simplifying give

$$\begin{aligned} \mathbb{E}u^\star = u_0 &- v\{\underline{h}[(1 - \pi_1)\bar{g} + \pi_1\hat{S}_1] + \bar{h}[(1 - \pi_2)\bar{g} + \pi_2\hat{S}_2]\} \\ &+ (1 - \underline{h}\pi_1 - \bar{h}\pi_2)A - \underline{h}d_1(\underline{\theta}) + \bar{h}d_2(\bar{\theta}). \end{aligned} \tag{20}$$

Finally, using (17) and (18) allows us to rewrite the conditions $0 \leq (1-y_2)I \leq I \leq L$ as $u^{-1}(u_0 - v) \leq u^{-1}(u_0) \leq W - P = u^{-1}(\hat{u}) = u^{-1}(u_0 + A - v\bar{g})$. As u is increasing, we must have

$$0 \leq v \leq A/\bar{g}. \tag{21}$$

[Figure 3 about here.]

The optimal solution is obtained by maximizing $\mathbb{E}u^\star$ given by (20) subject to (19) and (21). As illustrated in Figure 3, the insurer's break-even constraint $K(u_0, v) = \bar{W}$ corresponds to an increasing convex locus in the (u_0, v) plane. Iso-expected utility curves are increasing straight lines. Figure 3 illustrates the case of a corner solution at point E where $v = A/\bar{g}$, and thus with $I = L$. A calculation developed in the appendix shows that this is the case when $\bar{h}$ is large enough. More explicitly:

Proposition 3 *There exists $h^\star \in (0,1)$, such that there is full coverage (i.e., $I = L$) if $\bar{h} \geq h^\star$ and partial coverage (i.e., $I < L$) if $\bar{h} < h^\star$.*

Proposition 3 highlights the substitutability between the incentives provided by the law of insurance contracts and those provided by the indemnity schedule. Two regimes exist, according to whether $\overline{h}$ is larger or lower than the threshold $h^\star$.[20] If the proportion of individuals exerting only the low effort level is large (i.e. if $\bar{h} > h^\star$), it is socially optimal to offer an insurance contract with full coverage and to give the insurer the opportunity to cut the indemnity when the circumstances reveal a misconduct. On the contrary, if $\bar{h} < h^\star$, it is difficult for the insurer to credibly claim that the insured did misbehave (i.e., this is possible only for x very large), and incentives are better provided by offering a contractual indemnity lower than the loss whatever the circumstances. This corresponds to the case depicted Figure 4.

[Figure 4 about here.]

Very often, the law of insurance contracts is rather coarse in specifying the conditions that allow the insurer to renege on the indemnity payment: either the insurer must pay the complete indemnity because the circumstances are not sufficient to convince the court that the insured did misbehave, or the insurer pays nothing when he can credibly make such an allegation. Formally, this is equivalent to restrict the cut function $z(x)$ to either 0 or 1, hence, a reduced indemnity $(1 - y_2)I = 0$ and a corresponding utility level $u_1 = u(W - P - L)$. We now investigate the consequences of this restriction on the contractual indemnity.

Similarly to (19), the insurer break-even constraint can be written as $\bar{W} \geq K(u_0, v)$, but with an ex ante policyholder's expected wealth given by

$$\bar{W} = W - L(\underline{h}\pi_1\hat{G}_1 + \bar{h}\pi_2\hat{G}_2)$$

and

$$K(u_0, v) = (1 - \underline{h}\pi_1\hat{G}_1 - \bar{h}\pi_2\hat{G}_2)u^{-1}(u_0 + A - v\bar{g}) + (\underline{h}\pi_1\hat{G}_1 + \bar{h}\pi_2\hat{G}_2)u^{-1}(u_0),$$

which is a convex function of u_0 and v, satisfying $\partial K/\partial u_0 > 0$ and $\partial K/\partial v < 0$. We

[20] $h^\star$ depends on all the parameters of the problem, including $\overline{h}, \pi_1$ and π_2.

also have

$$\begin{aligned}\mathbb{E}u^{\star} &= \underline{h}\{(1-\pi_1)(A-\bar{g}v+u_0)+\pi_1[\hat{G}_1u_0+\hat{S}_1u(u^{-1}(A-\bar{g}v+u_0)-L)]-d_1(\underline{\theta})\}\\ &+\bar{h}\{(1-\pi_2)(A-\bar{g}v+u_0)+\pi_2[\hat{G}_2u_0+\hat{S}_2u(u^{-1}(A-\bar{g}v+u_0)-L)]-d_2(\bar{\theta})\}\\ &\equiv U(u_0,v)\end{aligned}$$

which is a concave function of u_0 and v if u is DARA, with $\partial U/\partial u_0 > 0$ and $\partial U/\partial v < 0$. Calculation shows that $v < A/\bar{g}$, i.e. $I < L$ at the optimum, as illustrated in Figure 5.

[Figure 5 about here.]

Proposition 4 *When $y_2 \in \{0,1\}$, the optimal contract provides partial coverage, i.e., $I < L$.*

Hence, when the law of insurance contract is restricted to an "all or nothing" type of decision, its efficiency is limited and it is necessary to reinforce the incentives provided by the terms of the contract by offering only a reduced coverage.

5 The burden of proof

In the two previous sections, we have assumed that the information x on the circumstances of the loss, known to the policyholder and possibly observed by the insurer through an audit, is hard evidence: this information could be presented as such to a court, without incurring any supplementary cost. Realities are often more complex. Auditing a claim may allow the insurer to examine all information about the circumstances of an accident, but persuading a judge that the policyholder misbehaved in a certain way frequently requires more. Put differently, it is one thing for the insurer to know the circumstances of an accident through an audit and make its own judgment, and quite another to be able to provide verifiable information to a court, if it is required to do so. To distinguish these two levels of information, we now consider that auditing a claim only provides soft information about the circumstances of the accident: the insurer can observe x through an audit, but more evidence is needed to sustain an allegation in court (i.e., to upgrade observed circumstances to verifiable hard information). Sustaining a claim by verifiable information is also costly to the policyholder, although this cost is presumably smaller than the investigation cost incurred by the insurer.[21] This raises the question of the allocation of the burden of proof, i.e. the obligation of a party to produce the evidence that will prove its claim against the other party who is given the "benefit of the doubt". These verification costs are at the origin of an additional efficiency loss by comparison with the third-best case where audit directly provides hard information on the circumstances of the accident. Accordingly, we qualify the optimal allocation in this setting as a fourth-best solution to the insurance moral hazard problem.

Consider first that the burden of proof is on the policyholder, and the insurer is given the benefit of the doubt. We assume that it costs k_P per dollar of indemnity to the insured to produce evidence on the circumstances of the loss. We also assume that evidence can only be found on the actual events that generated the loss, i.e. that the information cannot be distorted or forged.[22] Thereby, the judge cannot be misguided by the policyholder with some piece of fake evidence sustaining an information level different from x. In other words, the policyholder can only prove (at cost k_P) that the circumstances are x if she wants to refute the allegation of the insurer. The policyholder-insurer interaction is modified as follows:

- **Stages 1 to 3** are unaffected, but the information x obtained by the insurer in case of an audit is only soft information. Hence, the insurer's allegation $b \in \{2, \ldots, n\}$ is not sustained by evidence, i.e., it does not allow the judge to verify the true value x.

- **Stage 4:** the policyholder may decide to contest in court the insurer's allegation by transmitting verifiable information about x to the judge at cost k_P. In that case, the judge proves the insured right if $\hat{b}(x) < b$, which entitles the insured to receive the full indemnity I. Otherwise, the judge confirms the insurer's allegation and the rule of law y_b applies.

- **Stage 5** is unaffected.

Note that the contractual indemnity I is entirely paid if the judge proves the insured right. We will discuss this assumption below, but we may already consider it as corresponding to a bad faith clause frequently applied in insurance law: if $\hat{b}(x) < b$,

[21]For instance, the policyholder may need to yield reliable testimonies by witnesses or technical reports in corporate or personal liability issues, expert reports in property claims, medical certificates in medical malpractice claims or some kind of psychological report in work harassment cases.

[22]Crocker and Morgan (1997) develop a theory of claim falsification, in which policyholders spend resources to distort the information perceived by insurers about their loss.

the judge considers that the insurer was plainly deceptive and he obliges him to pay the contractual indemnity as bad faith penalty.[23]

After an audit that revealed x, the insurer knows that any allegation $b > \hat{b}(x)$ at stage 2 will be successfully contested by the insured at stage 4 if and only if $k_P \leq y_b$, because otherwise contesting the insurer's allegation would not be worthwhile. Let

$$x^\star = \inf\{x \in (0,1] : y_{\hat{b}(x)} \geq k_P\}. \tag{22}$$

When $x \geq x^\star$, it is optimal for the policyholder to contest any allegation $b > \hat{b}(x)$ with $y_b \geq y_{\hat{b}(x)}$. Indeed, bringing verifiable information about x to the court will persuade the judge that the allegation b is unfounded (i.e., cannot be justified on the balance of probabilities). Thus, the insurer will be condemned for bad faith and the policyholder proved right, with net gain for the latter

$$(y_b - k_P)I \geq (y_{\hat{b}(x)} - k_P)I \geq 0.$$

Consequently, when the insurer observes $x \geq x^\star$ through an audit, it is an optimal strategy for him to allege misbehavior $\hat{b}(x)$, and this will not be contested by the insured.

When $x < x^\star$, we have $y_{\hat{b}(x)} < k_P$ and the insurer knows that any allegation $b > \hat{b}(x)$ will be successfully contested by the policyholder if $y_b \geq k_P$, i.e., if $b \geq \hat{b}(x^\star)$. Hence, an optimal insurer's strategy consists in alleging the misconduct immediately lower than $\hat{b}(x^\star)$, i.e. $b = \hat{b}(x^\star) - 1$, which will not be contested by the policyholder.

The indemnity cut is

$$z(x) = \begin{cases} y_{\hat{b}(x^\star)-1} & \text{if } x < x^\star \\ y_{\hat{b}(x)} & \text{if } x \geq x^\star \end{cases} \tag{23}$$

and, whatever the case, the insured does not contest the insurer's allegation. The corresponding break-even condition for the insurer is given by

$$\begin{aligned} P \geq \int_\Theta \pi_{b^\star(\theta)} \Big\{ I + \int_0^1 cE[q(s)|x]\hat{g}_{b^\star(\theta)}(x)dx \\ - I \left(\int_0^{x^\star} y_{\hat{b}(x^\star)-1} + \int_{x^\star}^1 y_{\hat{b}(x)} \right) E[q(s)|x]\hat{g}_{b^\star(\theta)}(x)dx \Big\} dH(\theta). \end{aligned} \tag{24}$$

[23] Should the insurer have alleged $b = \hat{b}(x)$, then he would have been allowed to cut the indemnity by a fraction $y_{\hat{b}(x)}$.

Consider now that the burden of proof is on the insurer, and the policyholder is given the benefit of the doubt. When the insurer alleges some misconduct of the insured, he must provide verifiable information about x, which costs k_I per dollar of indemnity. The insured-insurer game after a claim is similar to the case where the information x is verifiable, but with the additional cost k_I for the insurer if he wants to allege some behavior $b \in \{2, ..., n\}$. Hence, after an audit at stage 2, the insurer alleges misconduct $b = \hat{b}(x)$ if and only if $y_{\hat{b}(x)} \geq k_I$, and thus the equilibrium indemnity cut is

$$z(x) = \begin{cases} 0 & \text{if } x < \tilde{x} \\ y_{\hat{b}(x)} & \text{if } x \geq \tilde{x} \end{cases} \tag{25}$$

where

$$\tilde{x} = \inf\{x \in (0, 1] : y_{\hat{b}(x)} \geq k_I\} \tag{26}$$

is the threshold above which the insurer finds it worthwhile to gather evidence on the circumstances of the claim, in order to convince the judge that the insured's behavior was $\hat{b}(x)$. The corresponding break-even condition is

$$P \geq \int_{\Theta} \pi_{b^\star(\theta)} \left\{ I + \int_0^1 cE[q(s)|x]\hat{g}_{b^\star(\theta)}(x)dx - I\int_{\tilde{x}}^1 (y_{\hat{b}(x)} - k_I)E[q(s)|x]\hat{g}_{b^\star(\theta)}(x)dx \right\} dH(\theta). \tag{27}$$

Comparing the two cases yields the following proposition.

Proposition 5 *If $k_P \leq k_I$ (or if $k_P - k_I$ is positive but not too large), an optimal fourth-best solution to the insurance moral hazard problem requires that the burden of proof is on the policyholder.*

The intuition of this result is twofold. Firstly, when the insurer has the burden of proof and wants to contest a claim, he must provide the court with costly evidence, whereas he only has to allege a misconduct that will not be contested when he has the benefit of doubt and the insured is given the burden of proof. Thus, attributing the burden of proof to the policyholder is a way to avoid the cost of transmitting hard evidence to the court. Secondly, it would be too costly for the insurer to contest a claim when x is small if he has the burden of proof, which implies the constraint $z(x) = 0$ if $x < \widetilde{x}$. Conversely, when the burden of the proof is given to the policyholder, the insurer has some leeway to nitpick when circumstances are favorable to the policyholder, since he may reduce the indemnity by a fraction $z(x) = y_{\hat{b}(x^\star)-1}$ if $x < x^\star$.

These are good reasons for the burden of proof to be given to the policyholder only if the cost of transmitting hard evidence is not larger (or, at least, not much larger) for the policyholder than for the insurer, i.e., if $k_P \leq k_I$ or at least if $k_P - k_I$ is positive but not too large. Should this not be the case, then the policyholder would hardly ever be in a position to contest any allegation of the insurer. In the extreme case where $k_I = 0$, then $\widetilde{x} = 0$ and the third-best and fourth-best optimal solutions coincide with the burden of the proof given to the insurer.

We have assumed that the insurer must pay the entire contractual indemnity I, and not the reduced indemnity $I[1 - y_{\hat{b}(x)}]$, if he has initially the benefit of the doubt but he is contradicted by the judge. Having to pay the additional amount $y_{\hat{b}(x)}I$ may thus be interpreted as a penalty reflecting the importance of utmost good faith in insurance law.[24] We could move further in that direction. An hypothetical legal regime would consist in giving the burden of the proof to the policyholder, with insurer bad faith penalty at least equal to $[k_P - (y_b - y_{\hat{b}(x)})]I$ to be paid to the policyholder, in addition to the indemnity $[1 - y_{\hat{b}(x)}]I$ when a misconduct $b > \hat{b}(x)$ is alleged by the insurer, and successfully contested by the policyholder. Under such a regime, the policyholder is incentivized to contest any exaggerated allegation about her behavior because her cost $k_P I$ will be reimbursed if the judge agrees with her, and consequently the insurer restricts himself to indemnity cuts that can be sustained in court. The third-best optimal solution could be reached in such an hypothetical legal regime, in which the policyholder has the burden of the proof and the insurer has to pay her a sufficiently high bad faith penalty. Unfortunately, although in some cases courts may condemn insurers to pay compensatory penalties to policyholders, it would be unwise to conclude that this is the ultimate solution to the moral hazard problem. In practice, courts cannot assess the policyholders' cost of gathering evidence on the circumstances of their loss. This is private information that cannot be easily estimated. Obliging the insurer to pay the contractual indemnity in case of bad faith is thus a convenient, albeit less effective, alternative to the cost-reimbursement solution.

[24]From an empirical standpoint, insurers' bad faith penalty play a more or less important role according to specificities of insurance law. In the US, most states recognize the right of policyholders to file private lawsuits against insurers alleging unfair claim settlement practices involving first-party insurance coverage. Some states consider bad faith as a contract breach, while others consider it as a tort, allowing the policyholder to recover for all harm or injuries sustained, including legal expenses, economic loss, and mental distress, while punitive damages may also be awarded. Tennyson & Warfel (2009) describe these various approaches. Asmat & Tennyson (2014) show that, on average, tort liability for insurer's bad faith is associated with higher settlement amounts.

6 Conclusion

Our results can be recapped in a few words. When the insurance policy does not specify the indemnity payment according to all the contingencies that may characterize the claim, insurance contracts are incomplete. In a context of moral hazard, conditioning the indemnity on the circumstances of the loss is nevertheless desirable in order to incentivize policyholders to exert an adequate level of effort. This can be indirectly reached through provisions of insurance law. Legal principles that allow insurers to reduce or cancel compensation under unfavorable claim circumstances act as an incentive device, and improve contractual efficiency. Their application is constrained by the judgment of courts who decide on the balance of probabilities. When audit only yields non-verifiable information to the insurer, then the burden of proof should be given to policyholders, while threatening bad faith penalties against insurers.

These conclusions have been reached in a setting that could be easily extended in several directions. In particular, we have assumed that courts decide on the balance of probabilities, because, in common law, this is the most usual standard of proof in case of civil suits. Although the "proof beyond reasonable doubt" standard is mainly used in criminal trial, it could be incorporated in our analysis by increasing from 1/2 to a higher value the probability threshold above which courts agrees with the party that has the burden of proof. The relationship between insurer, insured and courts could also be modelled in a richer and more interactive way, but the same fundamental trade-off between conditioning the indemnity on circumstances to incentivize policyholders and limiting the opportunism of the insurers would remain, and thus similar results would emerge.

Much remains to be done in order to explore the consequences of the incompleteness of insurance contracts in other directions. The issues of insurance fraud and bad faith in insurance contracting are of special interest. In particular, many theory papers about insurance fraud have restricted their attention to models where opportunistic policyholders file claims although they did not suffer any loss, or build up their claims above their true loss, and insurers may verify claims through costly audits.[25] In practice, such a clear-cut framework is far less frequent than more fuzzy situations where policyholders may claim they were in good faith because their situation was not clearly specified in their insurance policy, and ambiguity prevails on the duty of the contracting parties. Whatever the interaction process between policyholder and insurer, whether it be an amicable settlement or a litigation process, its outcome will be determined

[25]See Picard (2013) for a survey on the economic analysis of insurance fraud.

by stipulations of insurance law, such as the definition and consequences of misrepresentation and non-disclosure, the legal regime for bad faith in claims settlement, the interpretation of contractual exclusion clauses or the allocation of the burden of proof.[26]

It is also worth emphasizing that similar issues arise in other types of principal-agent relationships, where the incompleteness of contracts results from the difficulty to objectively measure individualized performance signals (e.g., customer satisfaction or involvment in cooperative tasks, in the case of job-related activities). The literature on relational contracts (Levin, 2003; MacLeod, 2003) characterizes mechanisms that may allow the truthful revelation of unverifiable information in such settings. For instance, the payment of bonuses may be transferred to a third-party acting as an independent referee, similar to a private or state-sponsored ombudsman. We have followed another route, by focusing attention on the role of contract law. Clearly, both approaches are complementary, as illustrated by insurance and labor markets, where the enforcement of legal principles and arbitration process coexist.

[26]As an illustration, see Tennyson & Warfel (2009) and Asmat & Tennyson (2014) on the effect of the insurance bad faith legal regime on claims settlement, the settlement process, and insurance markets.

References

Abraham, K. S. (1986). *Distributing risk: insurance, legal theory, and public policy.* New Haven: Yale University Press.

Asmat, D. P. & Tennyson, S. (2014). Does the threat of insurer liability for bad faith affect insurance settlements? *Journal of Risk and Insurance*, 81(1), 1–26.

Bourgeon, J.-M. & Picard, P. (2014). Fraudulent claims and nitpicky insurers. *The American Economic Review*, 104(9), 2900–2917.

Demougin, D. & Fluet, C. (2006). Preponderance of evidence. *European Economic Review*, 50(4), 963–976.

Demougin, D. & Fluet, C. (2008). Rules of proof and courts and incentives. *The RAND Journal of Economics*, 39(1), 20–40.

Fagart, M.-C. & Fluet, C. (2009). Liability insurance under the negligence rule. *The RAND Journal of Economics*, 40(3), 486–508.

Handel, B. R. & Kolstad, J. T. (2015). Health insurance for" humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8), 2449–2500.

Hart, O. & Moore, J. (1999). Foundations of incomplete contracts. *The Review of Economic Studies*, 66(1), 115–138.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(1), 79–91.

Levin, J. (2003). Relational incentive contracts. *American Economic Review*, 93(3), 835–857.

MacLeod, W. B. (2003). Optimal contracting with subjective evaluation. *American Economic Review*, 93(1), 216–240.

MacLeod, W. B. (2007). Reputations, relationships, and contract enforcement. *Journal of economic literature*, 45(3), 595–628.

Malcomson, J. M. (2012). Relational Incentive Contracts. In R. Gibbons & J. Roberts (Eds.), *The Handbook of Organizational Economics* chapter 2. Princeton University Press.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2), 380–391.

Picard, P. (2013). Economic analysis of insurance fraud. In G. Dionne (Ed.), *Handbook of Insurance*, volume 26 of *Huebner International Series on Risk, Insurance and Economic Security* (pp. 349–395). Springer.

Schwarcz, D. (2008). Redesigning consumer dispute resolution: A case study of the british and american approaches to insurance claims conflict. *Tul. L. Rev.*, 83, 735–812.

Scott, R. E. & Triantis, G. G. (2005). Incomplete contracts and the theory of contract design. *Case W. Res. L. Rev.*, 56(1), 187–201.

Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship. *The Bell Journal of Economics*, 10(1), 55–73.

Shavell, S. (1980). Damage measures for breach of contract. *The Bell Journal of Economics*, 11(2), 466–490.

Tennyson, S. & Warfel, W. J. (2009). Law and economics of first-party insurance bad faith liability. *Connecticut Insurance Law Journal*, 16(1), 203–242.

Thaler, R. H. & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.

Williamson, O. E. (1985). *Yhe Economic Institutions of Capitalism: Firms, markets, relational Contracting*. Free Press.

Works, B. (1998). Excusing nonoccurrence of insurance policy conditions in order to avoid disproportionate forfeiture: Claims-made formats as a test case. *Conn. Ins. LJ*, 5(2), 505–673.

Appendix

A Proof of Proposition 1

Let $\mu_b(\theta) \geq 0$ and $\lambda \geq 0$ be Kuhn-Tucker multipliers associated with (3) and (4), respectively. Denoting $W(\theta, x) \equiv W - P - L + I(\theta, x)$, the first-order optimality conditions w.r.t. $I(\theta, x)$ and P lead to

$$u'(W(\theta,x))\left[1-\sum_{b\neq b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\left(\frac{\pi_b}{\pi_{b^\star(\theta)}}\frac{\widehat{g}_b(x)}{\widehat{g}_{b^\star(\theta)}(x)}-1\right)\right]\begin{cases}\leq\lambda & \text{if } \ I(\theta,x)=0,\\ =\lambda & \text{if } \ 0<I(\theta,x)<L,\\ \geq\lambda & \text{if } \ I(\theta,x)=L,\end{cases} \tag{28}$$

for all $(\theta, x) \in \Theta \times [0, 1]$.

Note that, for all $\theta \in \Theta$, the optimal solution is such that $b > b^\star(\theta)$ for all $b \neq b^\star(\theta)$ such that (3) is binding,[27] and also that $\mu_b(\theta) = 0$ for all $b \neq b^\star(\theta)$ such that (3) is not binding. Let

$$\phi_b(x) \equiv \frac{\widehat{g}_{b+1}(x)}{\widehat{g}_b(x)} \text{ for all } b = 1, ..., n-1.$$

Hence, the l.h.s. of (28) may be written as

$$u'(W(\theta,x))\left[1-\sum_{b>b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\left(\frac{\pi_b}{\pi_{b^\star(\theta)}}\prod_{i=b^\star(\theta)}^{b-1}\phi_i(x)-1\right)\right],$$

and it is decreasing if $b^\star(\theta) < n$, because $\phi_b'(x) > 0$. Consider the case where $b^\star(\theta) < n$. We have

$$I_x'(\theta,x)=\frac{u'(W(\theta,x))^2}{u''(W(\theta,x))}\sum_{b>b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\frac{\pi_b}{\pi_{b^\star(\theta)}}\frac{d}{dx}\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(x)\right]<0,$$

if $0 < I(\theta, x) < L$. We have $I(\theta, 0) < L$ - and thus $I(\theta, x) < L$ for all x - if

$$u'(W-P)\left[1-\sum_{b>b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\left(\frac{\pi_b}{\pi_{b^\star(\theta)}}\frac{d}{dx}\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(0)\right]-1\right)\right]\leq\lambda,$$

[27]Indeed if there exists $b \in \mathcal{B}$ such that $b < b^*(\theta)$ and the incentive constraint (3) is binding, then replacing $b^*(\theta)$ by b would reduce the expected insurance cost - i.e., the right-hand-side of (4) - because $\pi_b < \pi_{b^*(\theta)}$, without changing the policyholder's expected utility. This would contradict the optimality of the solution.

and otherwise, we have $I(\theta,x)=L$ if $0\leq x\leq \underline{x}(\theta)$ and $I(\theta,x)<L$ if $x>\underline{x}(\theta)$, with $\underline{x}(\theta)>0$ defined by

$$u'(W-P)\left[1-\sum_{b>b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\left(\frac{\pi_b}{\pi_{b^\star(\theta)}}\frac{d}{dx}\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(\underline{x}(\theta))\right]-1\right)\right]=\lambda.$$

Similarly, we have $I(\theta,1)>0$ - and thus $I(\theta,x)>0$ for all x - if

$$u'(W-P-L)\left[1-\sum_{b>b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\left(\frac{\pi_b}{\pi_{b^\star(\theta)}}\frac{d}{dx}\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(1)\right]-1\right)\right]\geq\lambda,$$

and otherwise, we have $I(\theta,x)=0$ if $\overline{x}(\theta)\leq x\leq 1$ and $I(\theta,x)>0$ if $x<\overline{x}(\theta)$, with $\overline{x}(\theta)>0$ defined by

$$u'(W-P-L)\left[1-\sum_{b>b^\star(\theta)}\frac{\mu_b(\theta)}{h(\theta)}\left(\frac{\pi_b}{\pi_{b^\star(\theta)}}\frac{d}{dx}\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(\overline{x}(\theta))\right]-1\right)\right]=\lambda.$$

If $b^\star(\theta)=n$, then the l.h.s. of (28) is equal to $u'(W(\theta,x))$, which implies that $W(\theta,x)$ and $I(\theta,x)$ do not depend on x.

B Proof of Lemma 1

Let

$$\begin{aligned}\Phi(b_0,x)&=\sum\nolimits_{b=b_0}^{n}\Pr(b|x)-\sum\nolimits_{b=1}^{b_0-1}\Pr(b|x)\\&=\frac{\sum_{b=b_0}^{n}g_b(x)\int_{\Theta_b^\star}dH(\theta)-\sum_{b=1}^{b_0-1}g_b(x)\int_{\Theta_b^\star}dH(\theta)}{\sum_{b\in\mathcal{B}}g_b(x)\int_{\Theta_b^\star}dH(\theta)},\end{aligned}$$

for $b_0\geq 2$. When $\hat{b}(x)\geq 2$, we have

$$\hat{b}(x)=\sup\{b_0\in\mathcal{B}|\Phi(b_0,x)\geq 0\}.$$

Let $x' > x$. Using strict MLRP yields

$$\begin{aligned} g_b(x') &> g_b(x)\frac{g_{b_0}(x')}{g_{b_0}(x)} \quad \text{if} \quad b > b_0, \\ g_b(x') &< g_b(x)\frac{g_{b_0}(x')}{g_{b_0}(x)} \quad \text{if} \quad b < b_0. \end{aligned}$$

Hence, if $b_0 \geq 2, x' > x$, we have $\Phi(b_0, x') > 0$ if $\Phi(b_0, x) \geq 0$. We deduce $\hat{b}(x') \geq \hat{b}(x)$ if $\hat{b}(x) \geq 2$ and $x' > x$, which implies that $\hat{b}(x)$ is non-decreasing in $[0, 1]$. It is thus a step function that takes its values in $\mathcal{B}$.

C Proof of Lemma 2

An increase in s shifts the conditional probability distribution of x in the sense of strong FOSD. Furthermore, $y_b \leq y_{b+1}$ for all $b \in \{1, ..., n-1\}$ and $\hat{b}(x)$ is non-decreasing from Lemma 1. Consequently, $y_{\hat{b}(x)}$ is non-decreasing with x and $\mathbb{E}[y_{\hat{b}(x)}|s]$ is non-decreasing with s. Hence, $\mathbb{E}[y_{\hat{b}(x)}|s]I \geq c$ implies $\mathbb{E}[y_{\hat{b}(x)}|s']I \geq c$ if $s' > s$, which proves the Lemma.

D Proof of Proposition 2

Assume first $c = 0$. In that case, $q(s) = 1$ for all s, i.e., $s^\star = -\infty$. Let us restrict the set of feasible solutions to $y_b = y \geq 0$ for all $b \geq 2$. We have $z(x) = 0$ if $x < x_2$ and $z(x) = y$ if $x \geq x_2$ with x_2 the solution of $\hat{b}(x_2) = 2$, or equivalently

$$\sum\nolimits_{b=2}^{n} \Pr(\tilde{b} = b|x_2) = \Pr(\tilde{b} = 1|x_2),$$

where $\tilde{b}$ is the behavior of an individual who is randomly drawn among the claimants. This condition can be written as

$$\sum\nolimits_{b=2}^{n} \frac{g_b(x_2)}{g_1(x_2)} \int_{\Theta_b^\star} dH(\theta) = \int_{\Theta_1^\star} dH(\theta). \tag{29}$$

The sets $\Theta_1^\star, \ldots, \Theta_n^\star$ depend on I, P and y, and thus (29) implicitly defines function $x_2(I, P, y)$, with $x_2(I, P, 0) = x_{2na}$ if I, P is the optimal no-audit contract. Let $\lambda > 0$ and $\mu_b(\theta) \geq 0$ for $b \in \mathcal{B}, \theta \in \Theta$, be Lagrange multipliers corresponding to the insurer's break-even constraint and the incentive constraints respectively. Denote $u(1)$ and $u'(1)$ (resp. $u(2)$ and $u'(2)$) the value of the utility function and of its derivative when $x < x_2$

(resp. when $x \geq x_2$). The first-order optimality conditions w.r.t. I and y are written as

$$\begin{aligned}
&\int_{\Theta} \left\{ \left[u'(1)G_{b^\star(\theta)}(x_2) + u'(2)(1-y)[1-G_{b^\star(\theta)}(x_2)] \right.\right.\\
&\left.\left. -\lambda \left[1 - y[1-G_{b^\star(\theta)}(x_2)]\right]\right\} dH(\theta) \right.\\
&- \int_{\Theta} \{ \textstyle\sum_{b\in\mathcal{B}} \mu_b(\theta) \left\{u'(1)[G_b(x_2) - G_{b^\star(\theta)}(x_2)]\right.\\
&\left. +u'(2)(1-y)[G_{b^\star(\theta)}(x_2) - G_b(x_2)]\right\}\} dH(\theta)\\
&- \frac{\partial x_2}{\partial I} \times A\\
&\geq 0, = 0 \text{ if } I < L,
\end{aligned} \tag{30}$$

and

$$\begin{aligned}
&I[\lambda - u'(2)] \int_{\Theta} [1 - G_{b^\star(\theta)}(x_2)] dH(\theta)\\
&+ u'(2)I \int_{\Theta} \textstyle\sum_{b\in\mathcal{B}} \mu_b(\theta)[G_{b^\star(\theta)}(x_2) - G_b(x_2)] dH(\theta)\\
&- \frac{\partial x_2}{\partial y} \times A\\
&\leq 0, = 0 \text{ if } y > 0,
\end{aligned} \tag{31}$$

where

$$\begin{aligned}
A = {}&[u(1) - u(2)] \int_{\Theta} \{g_{b^\star(\theta)}(x_2) + \textstyle\sum_{b\in\mathcal{B}} \mu_b(\theta)[g_{b^\star(\theta)}(x_2) - g_b(x_2)]\} dH(\theta)\\
&+ \lambda y \int_{\Theta} g_{b^\star(\theta)}(x_2) dH(\theta).
\end{aligned}$$

Suppose that $y = 0$ at an optimal solution. This implies $I < L$ and $x_2 = x_{2na}$, because $I = L$ would imply $b^\star_{na}(\theta) = n$ almost everywhere in Θ, which has been excluded by assumption. In that case, we have $u(1) = u(2) \equiv u, u'(1) = u'(2) \equiv u'$ and $A = 0$. (30) and (31) simplify to

$$\int_{\Theta} \{(u' - \lambda)\pi_{b^\star_{na}(\theta)} - u' \textstyle\sum_{b\in\mathcal{B}} \mu_b(\theta)(\pi_b - \pi_{b^\star_{na}(\theta)})\} dH(\theta) = 0, \tag{32}$$

and

$$
\begin{aligned}
&(\lambda - u') \int_\Theta [1 - G_{b^\star_{na}(\theta)}(x_{2na})] dH(\theta) \\
&+ u' \int_\Theta \sum\nolimits_{b\in\mathcal{B}} \mu_b(\theta)[G_{b^\star_{na}(\theta)}(x_{2na}) - G_b(x_{2na})] dH(\theta) \leq 0,
\end{aligned} \tag{33}
$$

respectively. Substituting the value of $u' - \lambda$ given by (32) into (33) yields

$$
\begin{aligned}
&\int_\Theta \sum\nolimits_{b\in\mathcal{B}} \mu_b(\theta)(\pi_b - \pi_{b^\star_{na}(\theta)}) dH(\theta) \times \int_\Theta [1 - G_{b^\star_{na}(\theta)}(x_{2na})] dH(\theta) \\
&+ \int_\Theta \pi_{b^\star_{na}(\theta)} dH(\theta) \times \int_\Theta \sum\nolimits_{b\in\mathcal{B}} \mu_b(\theta)[G_{b^\star_{na}(\theta)}(x_{2na}) - G_b(x_{2na})] dH(\theta) \leq 0.
\end{aligned} \tag{34}
$$

There exists $b \in \mathcal{B}$ and a positive-measure subset of Θ in which $\mu_b(\theta) > 0$, once again because otherwise we would have $b^\star_{na}(\theta) = n$ almost everywhere in Θ. Furthermore, we have $b > b^\star_{na}(\theta)$ and $\pi_b \geq \pi_{b^\star_{na}(\theta)}$ if $\mu_b(\theta) > 0$, which implies that the first product in (34) is positive. Similarly, MLRP implies FOSD, which yields $G_b(x_{2na}) < G_{b^\star_{na}(\theta)}(x_{2na})$. Since this is true for all $\theta \in \Theta$, the second product in (34) is also (weakly) positive, hence a contradiction.

Consequently, when $c = 0$ we have $z(x) > 0$ in a positive-measure subset of $[0, 1]$. As the optimal expected utility of the policyholder varies continuously with c, the previous conclusion remains true with $s^\star < +\infty$ when c is not too large.

E Proof of Proposition 3

There is a corner solution with $v = A/\bar{g}$ when the slope of the iso-utility lines is larger than the one of the $K(u_0, v) = \bar{W}$ at that point. We have

$$
\frac{\partial K(u_0, v)}{\partial u_0} = \frac{\underline{h}\pi_1 \hat{G}_1 + \bar{h}\pi_2 \hat{G}_2}{u'(u^{-1}(u_0))} + \frac{\underline{h}\pi_1 \hat{S}_1 + \bar{h}\pi_2 \hat{S}_2}{u'(u^{-1}(u_0 - v))} + \frac{1 - \underline{h}\pi_1 - \bar{h}\pi_2}{u'(u^{-1}(u_0 + A - v\bar{g}))}
$$

which gives

$$
\left.\frac{\partial K(u_0, v)}{\partial u_0}\right|_{v=A/\bar{g}} = \frac{1 - \hat{S}_1 \pi_1 \underline{h} - \hat{S}_2 \bar{h}\pi_2}{u'(u^{-1}(u_0))} + \frac{\underline{h}\pi_1 \hat{S}_1 + \bar{h}\pi_2 \hat{S}_2}{u'(u^{-1}(u_0 - v))},
$$

and

$$
\frac{\partial K(u_0, v)}{\partial v} = -\frac{\underline{h}\pi_1 \hat{S}_1 + \bar{h}\pi_2 \hat{S}_2}{u'(u^{-1}(u_0 - v))} - \bar{g}\frac{1 - \underline{h}\pi_1 - \bar{h}\pi_2}{u'(u^{-1}(u_0 + A - v\bar{g}))},
$$

Hence, the slope of the insurer's break-even locus when $v = A/\bar{g}$ is given by

$$\begin{aligned}
\left.\frac{dv}{du_0}\right|_{\substack{K=\bar{W} \\ v=A/\bar{g}}} &= \frac{(1-\underline{h}\pi_1\hat{S}_1 - \bar{h}\pi_2\hat{S}_2)/u'(u^{-1}(u_0)) + (\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)/u'(u^{-1}(u_0-v))}{\bar{g}(1-\underline{h}\pi_1 - \bar{h}\pi_2)/u'(u^{-1}(u_0)) + (\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)/u'(u^{-1}(u_0-v))} \\
&= \frac{1-\underline{h}\pi_1\hat{S}_1 - \bar{h}\pi_2\hat{S}_2 + (\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)u'(u^{-1}(u_0))/u'(u^{-1}(u_0-v))}{\bar{g}(1-\underline{h}\pi_1 - \bar{h}\pi_2) + (\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)u'(u^{-1}(u_0))/u'(u^{-1}(u_0-v))}.
\end{aligned}$$

Similarly, the slope of the iso-utility lines is given by

$$\left.\frac{dv}{du_0}\right|_{Eu^\star=cst} = \frac{1}{\bar{g}(1-\underline{h}\pi_1 - \bar{h}\pi_2) + \underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2}.$$

Defining

$$f(\gamma) \equiv \frac{1-\underline{h}\pi_1\hat{S}_1 - \bar{h}\pi_2\hat{S}_2 + \gamma(\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)}{\bar{g}(1-\underline{h}\pi_1 - \bar{h}\pi_2) + \gamma(\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)},$$

we have $(dv/du_0)|_{K=\bar{W};v=A/\bar{g}} = f(u'(u^{-1}(u_0))/u'(u^{-1}(u_0-v)))$ where $u'(u^{-1}(u_0))/u'(u^{-1}(u_0-v)) < 1$, while $(dv/du_0)|_{Eu^\star=cst} = f(1)$. We thus have $(dv/du_0)|_{K=\bar{W};v=A/\bar{g}} < (dv/du_0)|_{Eu^\star=cst}$ if f is increasing, i.e. if

$$\begin{aligned}
\bar{g}(1-\underline{h}\pi_1 - \bar{h}\pi_2) > 1-\underline{h}\pi_1\hat{S}_1 - \bar{h}\pi_2\hat{S}_2 &= 1-\pi_1\hat{S}_1 - \bar{h}(\pi_2\hat{S}_2 - \pi_1\hat{S}_1) \\
&= 1-\pi_1\hat{S}_1 - \bar{h}(\pi_2 - \pi_1)\bar{g},
\end{aligned}$$

hence if $\bar{g} > (1-\pi_1\hat{S}_1)/(1-\pi_1)$. Substracting 1 on both sides, this condition becomes

$$\bar{g} - 1 = \frac{\pi_1\hat{G}_1 - \pi_2\hat{G}_2}{\pi_2 - \pi_1} > \frac{1-\pi_1\hat{S}_1}{1-\pi_1} - 1 = \frac{\pi_1\hat{G}_1}{1-\pi_1},$$

or, reorganizing terms,

$$\begin{aligned}
0 &< (\pi_1\hat{G}_1 - \pi_2\hat{G}_2)(1-\pi_1) - (\pi_2-\pi_1)\pi_1\hat{G}_1 \\
&= \hat{G}_1\pi_1(1-\pi_2) - \hat{G}_2\pi_2(1-\pi_1) \\
&= G_1(x_2)G_2(0) - G_2(x_2)G_1(0),
\end{aligned}$$

hence

$$\frac{G_1(x_2)}{G_2(x_2)} > \frac{G_1(0)}{G_2(0)}.$$

Under this condition, the optimal solution satisfies $v = A/\bar{g}$, i.e. $I = L$, and as $A/\bar{g} > 0$, $J = (1-y_2)I < I$: the optimal solution involves contractual full insurance and the law of contracts allows the insurer to reduce the indemnity by an amount

$I - J = u^{-1}(u_0^*) - u^{-1}(u_0^* - A/\bar{g})$ where u_0^* is given by $K(u_0^*, A/\bar{g}) = \bar{W}$.

This condition can also be written as

$$\frac{\hat{G}_2(x_2)}{\hat{G}_1(x_2)} < \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} \equiv R < 1 \tag{35}$$

where $x_2 > 0$ is defined by $\Pr\{b = 2|x_2\} = 1/2$, i.e.

$$\frac{\hat{g}_2(x_2)}{\hat{g}_1(x_2)} = \frac{\pi_1(1-\bar{h})}{\pi_2\bar{h}}. \tag{36}$$

We know from MLRP, that $\hat{g}_2(x)/\hat{g}_1(x)$ is increasing for all $x \in (0,1]$. Using $\hat{g}_2(x)/\hat{g}_1(x) \to 0$ when $x \to 0$ and $\hat{g}_2(x)/\hat{g}_1(x) \to +\infty$ when $x \to 1$, implies $0 < x_2 < 1$. Given π_1 and π_2, (36) defines x_2 as a decreasing function of $\bar{h} : x_2 = x_2(\bar{h})$ with $dx_2/d\bar{h} < 0$, $x_2(\bar{h}) \to 1$ when $\bar{h} \to 0$ and $x_2(\bar{h}) \to 0$ when $\bar{h} \to 1$. Define $\psi(x) = \hat{G}_2(x)/\hat{G}_1(x)$ where $x > 0$. We have

$$\begin{aligned}\psi'(x) &= [\hat{g}_2(x)\hat{G}_1(x) - \hat{g}_1(x)\hat{G}_2(x)]/\hat{G}_1(x)^2 \\ &= \hat{G}_1(x)^{-2}\int_0^x [\hat{g}_2(x)\hat{g}_1(y) - \hat{g}_1(x)\hat{g}_2(y)]dy \\ &> 0\end{aligned}$$

since, from MLRP, $\hat{g}_2(x)/\hat{g}_1(x) > \hat{g}_2(y)/\hat{g}_1(y)$ for $y < x$. Moreover, $\lim_{x\to 0}\psi(x) = \lim_{x\to 0}\hat{g}_2(x)/\hat{g}_1(x) = 0$ using l'Hôpital's rule, and $\psi(1) = 1$. Function $\psi(x)$ is thus an increasing bijection from $(0,1]$ to itself. Condition (29) can be written as $\psi(x_2(\bar{h})) < R$, hence $x_2(\bar{h}) < \psi^{-1}(R)$, i.e. $\bar{h} > x_2^{-1}(\psi^{-1}(R)) \equiv h^\star$. We thus have $I = L$ if $\bar{h} \geq h^\star$ and $I < L$ if $\bar{h} < h^\star$.

F Proof of Proposition 4

Using $y_2 = 1$ gives

$$\left.\frac{dv}{du_0}\right|_{\substack{K=\bar{W}\\ v=A/\bar{g}}} = \frac{1}{\bar{g}(1 - \underline{h}\pi_1\hat{G}_1 - \bar{h}\pi_2\hat{G}_2)} = \eta$$

and

$$\left.\frac{dv}{du_0}\right|_{Eu^\star = cst} = \Gamma\left(\frac{u'(u^{-1}(A - \bar{g}v + u_0) - L)}{u'(u^{-1}(A - \bar{g}v + u_0))}\right)$$

where function $\Gamma(\cdot)$ is given by

$$\Gamma(\gamma) = \frac{1 + \underline{h}\pi_1\hat{S}_1(\gamma - 1) + \bar{h}\pi_2\hat{S}_2(\gamma - 1)}{\bar{g}[1 + \underline{h}\pi_1(\hat{S}_1\gamma - 1) + \bar{h}\pi_2(\hat{S}_2\gamma - 1)]}$$

with $\Gamma(1) = \eta$ and

$$\Gamma'(\gamma) = -\frac{(\underline{h}\pi_1\hat{S}_1 + \bar{h}\pi_2\hat{S}_2)(\underline{h}\pi_1\hat{G}_1 + \bar{h}\pi_2\hat{G}_2)}{\bar{g}[1 + \underline{h}\pi_1(\hat{S}_1B - 1) + \bar{h}\pi_2(\hat{S}_2B - 1)]^2} < 0.$$

As $u'(u^{-1}(u_0)-L) > u'(u^{-1}(u_0))$, we have $(dv/du_0)|_{Eu^\star=cst,v=A/\bar{g}} < (dv/du_0)|_{K=\bar{W},v=A/\bar{g}}$, which implies that $v < A/\bar{g}$, i.e. $I < L$ at the optimum.

G Proof of proposition 5

In the fourth-best world, an allocation a is defined by a contract (P, I), an audit strategy $q(\cdot)$ with threshold $s^\star$, insurance law rules $\{y_b, b \in \mathcal{B}\}$, a behavioral function $b^\star(\cdot)$ and a (potential) allegation function $\hat{b}(\cdot)$. Let a_I be the optimal allocation when the burden of proof is given to the insurer, with verification cost k_I. It satisfies (6)-(10) and (12)-(14) together with (26), and condition (27) which is binding. As $z(x) = 0$ for $x < \tilde{x} = \inf\{x \in (0, 1] : y_{\hat{b}(x)} \geq k_I\}$, we may assume $y_{\hat{b}(x)} = 0$ if $x < \tilde{x}$ wlg.

Let $\mathcal{A}_P$ the set of feasible allocations when the burden of proof is given to the policyholder, with cost of verifiable information k_P. An allocation a belongs to $\mathcal{A}_P$ if it satisfies (6)-(10) and (12)-(14), together with (22) and (24). Suppose that $k_P = k_I$. We have $x^\star = \inf\{x \in (0, 1] : y_{\hat{b}(x)} \geq k_P\} = \tilde{x}$, the indemnity cut functions (23) and (25) are the same, and since a_I satisfies (27) as an equality, it also satisfies (24) strictly. Hence $a_I \in \mathcal{A}_P$. However, since (24) is satisfied but not binding, a_I is not the optimal allocation when the burden of the proof is given to the policyholder, i.e., there exists $a \in \mathcal{A}_P$ that leads to a policyholder's expected utility higher than with a_I. Since the optimal expected utility is non-increasing w.r.t. k_P when the policyholder has the burden of the proof, the same conclusion holds when $k_P < k_I$, and it remains true if $k_P - k_I$ is positive but not too large.

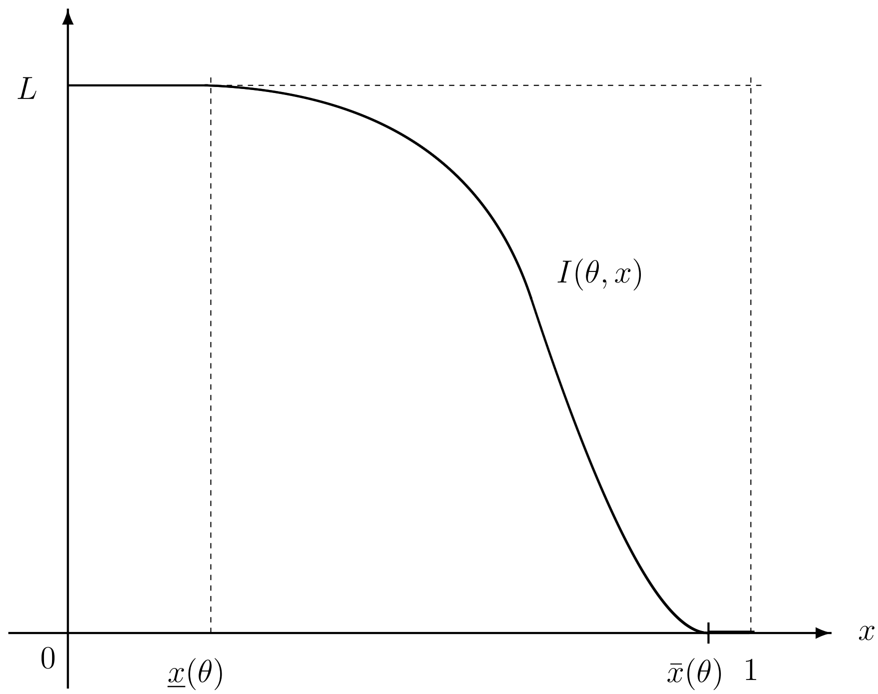

Figure 1: Optimal indemnity

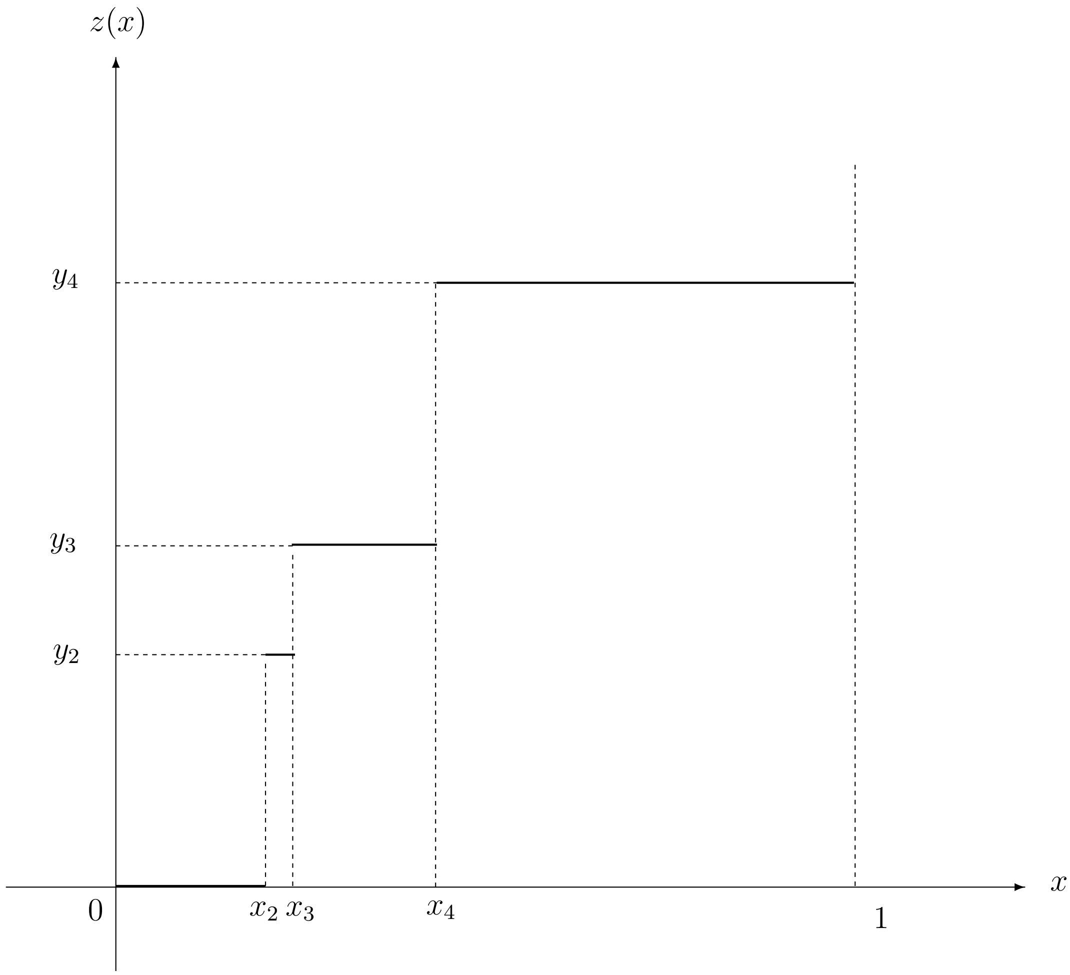

Figure 2: Optimal indemnity and the balance of probability.

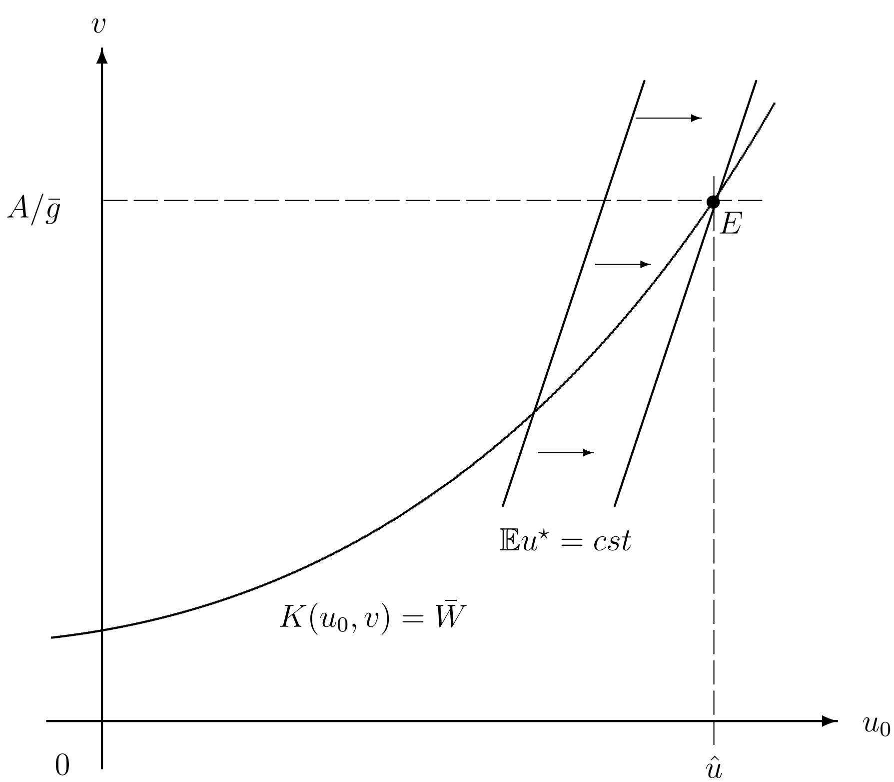

Figure 3: Optimal contract when $\bar{h} > h^{\star}$

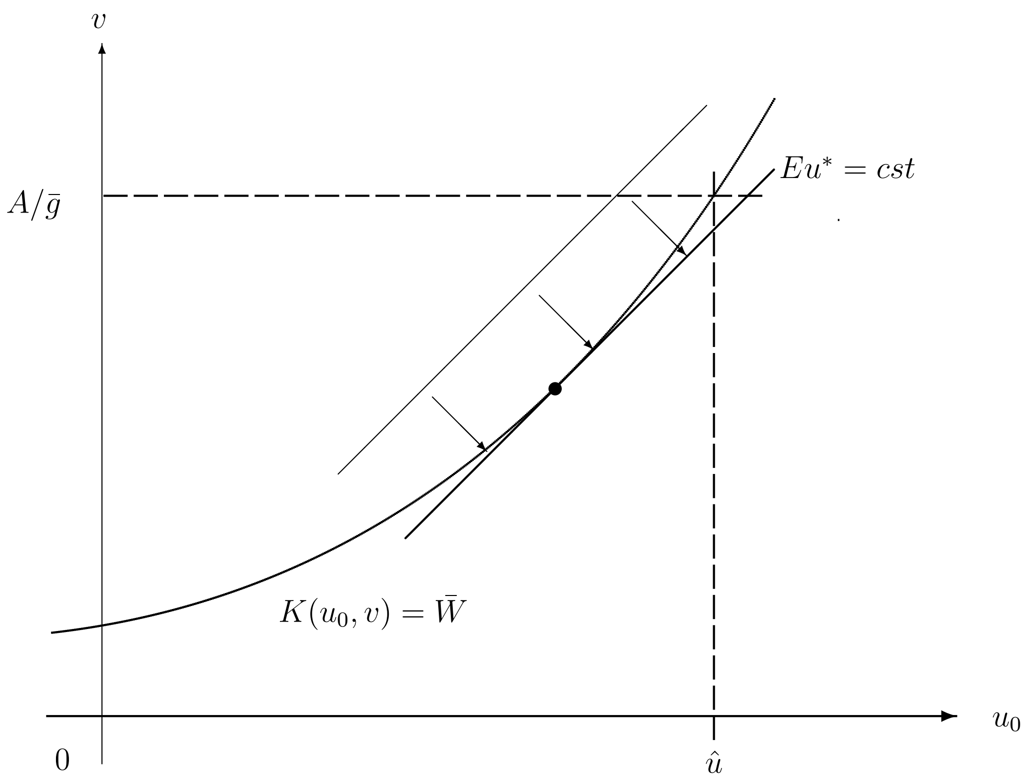

Figure 4: Optimal contract when $\bar{h} < h^\star$

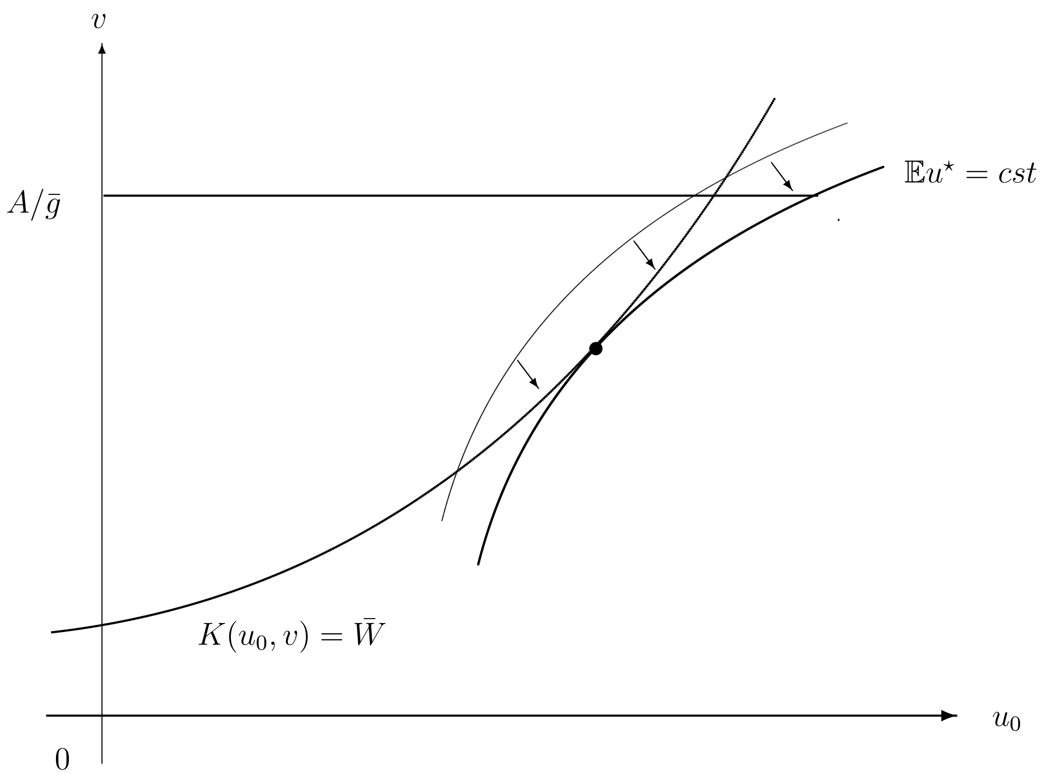

Figure 5: Optimal constrained contract