



HAL
open science

Insurance Law and Incomplete Contracts

Jean-Marc Bourgeon, Pierre Picard

► **To cite this version:**

| Jean-Marc Bourgeon, Pierre Picard. Insurance Law and Incomplete Contracts. 2019. hal-01830360v3

HAL Id: hal-01830360

<https://hal.science/hal-01830360v3>

Preprint submitted on 13 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Insurance Law and Incomplete Contracts

Jean-Marc Bourgeon* Pierre Picard†

Revised September 27, 2019

Abstract

Under moral hazard most insurance contracts condition the coverage neither on the contingencies under which policyholders have chosen their unobserved behavior, nor on the concrete features of the event that generated the damages. 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 reduce or cancel 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 with 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 a propensity to challenge the legitimacy of claims by invoking provisions or the practices of insurance law.

From a risk-sharing standpoint, this unpredictability of coverage weakens the efficiency of insurance contracting.¹ 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 detail in the insurance contract, and it is often limited to exclusions or force majeure clauses.² In other words, more often than not, insurance contracts are incomplete, and the resulting ambiguity may introduce a conflicted dimension into the insurer-insured relationship.³

This incompleteness can be attributed to the insurer's underwriting cost and to the

¹The efficiency loss due to the unpredictability of insurance coverage has been highlighted in the literature on insurance contract non-performance, whose canonical formal set-up was introduced by Doherty & Schlesinger (1990). Insurance non-performance occurs when the insurer is insolvent, but also when he denies coverage in an abusive manner. See Harrison & Ng (2018) for a behavioral evaluation of the welfare effects of non-performance risk.

²For 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 the necessity to call on affiliated service providers (e.g., managed care networks), except in the case of force majeure. However, insurance contracts rarely include conditional indemnities, i.e. 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.

³The incompleteness of insurance contracts is subject to much attention from lawyers, judges and policymakers. 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 insurers 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).

policyholder's cognitive cost,⁴ both of them being prohibitive in the case of complete 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. However, the incompleteness of insurance contracts does not necessarily break 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. Alleging some misconduct by the policyholder then appears as an indirect way to condition the insurance coverage on the circumstances of the loss, provided, of course, that the judge agrees with the insurer.

The concern about moral hazard holds a significant place in the law of insurance contracts. It is viewed by law makers and judges as a situation where the policyholder's negligence or more severe misconducts breach the probabilistic link between a peril faced by the policyholder under normal behavior (e.g., "reasonable care" in the case of property insurance) and operative events leading to losses. Beyond contractual stipulations aiming at curbing moral hazard (e.g., deductible, co-payment or experience rating), this breach requires that negligence, deliberate recklessness or willful misconducts are, in legal terms, subject to "contract exception." In other words, they are considered as reasons for which the compensation may be reduced or refused, on grounds of the insured's behavior alleged by the insurer. In the light of the circumstances of the event, the judge - if the case ends up in the courts - has to decide whether such an exception applies. This goes through a case by case approach, in which the circumstances of the loss are examined and the judge has to decide whether or not, on the balance of probabilities, the allegedly inadequate behavior of the policyholder motivates an exception for negligence.⁵

⁴For 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 hassle costs.

⁵Under Common Law, exceptions of negligence have been construed rather restrictively, with the objective of not constituting legal obstacles to the development of insurance markets, particularly through opportunistic insurers' behaviors. In the history of insurance law, this restrictive approach is associated with Lord Diplock, from the name of the member of the House of Lords who delivered the judgment in *Fraser vs Furman Ltd* (1967), in which this principle was stated (see Clarke, 1997, section 19.2).

Insurance lines	Misconducts	Legal means
Health, life or credit insurance	Misrepresenting pre-existing health conditions. Non-disclosing previous health events, hiding smoker status.	Misrepresentation Non-disclosure
Homeowner insurance	Not taking maintenance measures if a major storm is announced.	Duty of care
Theft insurance	Non revealing valuables that could be stolen. Undervaluing an asset.	Non-disclosure Misrepresentation
Automobile insurance	Driving on worn tires. Driving under the influence of alcohol.	Duty of care Reckless conduct
D&O liability insurance	Lack of supervision by corporate board. Financial statement falsification.	Duty of care Misrepresentation

Table 1: Examples of misconducts and legal means.

In the practice of insurance law, there are a number of legal means that can be invoked by insurers, either before a court or in an out-of-court settlement, to motivate a decrease in indemnity or to back the pure and simple denial of the claim. They are a great many misconducts that may occur, potentially related to a small number of specific legal means, with combinations and variable intensities in the application of each of them. Table 1 provides a few examples of such a correspondence. Legal means, like “duty of care” or “reckless conduct”, are directly related to willful actions at the origin of the insured’s loss. Means like “misrepresentation” or “non-disclosure” may be used when the policyholder has deliberately falsified his/her risk exposure, hence hiding his actual behavior and contradicting the principle of “utmost good faith” which is at the heart of insurance law.

In a case such as automobile insurance, the operative events leading to losses are usually well documented through friendly reports, complemented by expert and police reports when necessary, and insurance by-laws assign responsibilities to the parties without much ambiguity. Furthermore, available information repeatedly accumulated by insurers allows them to offer sophisticated contracts, including provisions such as pay-as-you-drive and experience-rating, that contribute to reduce the intensity of moral hazard. This will be reinforced by the sharp development of telematics allowing insurers (or firms in the case of fleet insurance) to closely monitor the behavior of drivers. This is expected to reduce even more the moral hazard problem. However, for many other

property-casualty and liability insurance risks, the circumstances of the losses are much less objectified and transparent, and conflicts of interpretation abound, with insurance law as a last resort.

As a typical example of how contract law interferes with the settlement of insurance claims under moral hazard, consider the case of Directors and Officers liability insurance (D&O). The question as to whether D&O insurance weakens the deterrence role of security law protecting minority shareholders has been extensively studied. Although such analyses are made difficult by the lack of data transparency, there is some evidence that D&O insurance creates corporate moral hazard through misconducts of managers and directors.⁶ Insurers try to reduce this moral hazard by limiting their coverage through deductibles, coinsurance and upper limits. However, the incentive effects of such provisions remain obviously limited, and relying on insurance law is often the only solution available to insurers. In this respect, the Enron bankruptcy is an almost caricatural textbook case. In November 2001, Enron recognized that its financial statements for 1997 through the first half of 2001, included in its SEC filings, were wrong and would have to be restated. Following this announcement, several of Enron's D&O insurers, and first of all AEGIS, its Bermuda mutual insurer, filed motions with the Bankruptcy Court asserting that, because of such a material misrepresentation, they were not bound by their policies and no indemnity should be paid. Enron, World-Com, and many other minor cases have been at the origin of endless debates before the courts, in which the good or bad faith of directors was a crucial element for courts to decide whether D&O insurance contracts were fully or partially rescinded, hereby inducing dramatic consequences for managers and directors accused of wrongdoings.

In cases that set precedent in several jurisdictions, the courts approved the insurer's indemnity denial by referring to the moral hazard problem, for the first time advocated in explicit terms by Judge Richard Posner.⁷ The courts tend to approve the denial of coverage when the insured did not comply with pre-existing obligations (such as claims for unpaid wages to workers or benefits to pensioners), when he/she gained a personal profit or other advantage to which he/she was not legally entitled, or in the case of the breach of a written contract by the insured,⁸ to name but a few examples

⁶See Ambrose et al. (2013) for a survey. See Baker & Griffith (2006) and Zou et al. (2008) on publicly traded firms in the US and in China, respectively.

⁷Posner's moral hazard argument was stated in *May Department Stores Co. vs Federal Insurance Co* (2002).

⁸As explained by Judge Posner in the decision about *Krueger International Inc vs Royal Indemnity Co*. (2007), "Moral hazard provides a further explanation for the distinction that the policy makes between written and oral contracts. The breach of a written contract will often be a deliberate act by the insured, while the breach of a contract created by an oral representation of an employee is likely

of loss circumstances affecting the court decisions. Most of the time, the question is whether the complaint from shareholders, workers or customers results from errors or occasional carelessness inherent within the framework of business life or from the willful misconduct of D&Os.

In what follows, we will investigate the issue of moral hazard under incomplete insurance contracts from a theoretical standpoint. To do so, we will be considering 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 transfer their risk exposure entirely to risk-neutral insurers. The variety of situations 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 concerns 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. it is impossible to spell out the relationship between the indemnity reduction and the many states and circumstances of loss. In this third-best setting, the insurer may decide to audit a claim in order to obtain evidence on 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. In the case of judicial recourse, we assume that the judge has perfect knowledge of the distribution of the insured behaviors in the population, and that his decision is made according to the "balance of probabilities" or to a higher standard of proof.⁹ This corresponds to a judicial system that allows insurers to reduce or cancel indemnities

to be, from the insured's standpoint, an unavoidable accident. The difference lies in the nature of the act that precipitates the breach: a deliberate decision by the insured, on the one hand, and the careless or unauthorized act of an employee on the other."

⁹In Common Law, proof based on the balance of probabilities is the standard in use in cases of civil suits, while proof beyond reasonable doubt is adopted in criminal cases. Demougin & Fluet (2006) show that the balance of probabilities decision rule provides maximal incentives to exert care. See also Demougin & Fluet (2008) for an analysis of the case of imperfect evidence.

in circumstances that are likely to reveal severe misconduct by the policyholder. This leeway provided by the 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 into contractual clauses. We will also consider the particular case of a “standard of negligence” whereby the insurer is allowed to fully deny the coverage above a negligence threshold, and has to pay the contractual indemnity otherwise, the judge’s approval being based on a standard of proof.

Finally, we will investigate the burden of proof issue, when the costs of providing verifiable information (i.e. strong evidence that may be 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 of avoiding the cost of transmitting hard evidence to the court, because the insurer has only to allege misconduct that will not be contested at equilibrium. This result is robust to the introduction of out-of-court-settlements, but it requires that the insurer faces a bad faith penalty if he is contradicted by the judge. In other words, giving the burden of 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 sources. The first 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 in gauging and verifying the many aspects of performance such as teamwork or initiative.¹⁰ 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. This 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.¹¹ 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

¹⁰See MacLeod (2007), or Malcomson (2012), for an overview of the “relational contracts” literature.

¹¹See for instance Shavell (1980) on damage measures for breach of contract, and Scott & Triantis (2005), who challenge 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. See Spier (2007) for a survey of the academic literature on the economics of litigation.

discretionary power. The analysis of the negligence rule by Fagart & Fluet (2009) is illustrative of this approach in the case of liability insurance.¹² A connection may also be made with the analysis of insurance fraud by Bourgeon & Picard (2014): by allowing insurers to reduce the indemnity according to the circumstances of the loss, insurance law acts as an incentive device to perform costly audits. Burden of proof rules, which require a specified party to produce evidence on a contested issue, are central to the adversary system. We highlight conditions under which the burden of proof should be on the policyholders, provided that insurers are threatened with bad faith penalties.

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, resulting in a distribution of effort levels 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 concerns 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. We still consider in this case that 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 basis of a standard of proof, such as the balance of probabilities. Section 4 focuses attention on simplified cases, with only two or three types of behavior. This allows us to highlight the trade-off between incentives through the rules of law and through partial insurance coverage. Section 5 considers the case where auditing claims only provides soft information about the circumstances of the loss, i.e. it only 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.

¹²Fagart & 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 decisions {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 a competitive insurance market where insurers provide coverage to risk-averse individuals (households or firms) against accidents that may result in a loss L . The occurrence of the loss depends on the policyholder's behavior which is indexed by $b \in \mathcal{B} = \{1, 2, \dots, n\}$ and ranks the probability of accident π_b increasingly, i.e. $\pi_1 < \pi_2 < \dots < \pi_n$. Hence, $b = 1$ corresponds to a cautious behavior with the lowest probability of accident, and the other behaviors $b \in \{2, \dots, n\}$ refer to various types of misconduct, 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, which we refer to as the “state”, when she chooses her behavior. We assume $\theta \in \Theta$, where Θ is the (finite or infinite) 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.¹³

By way of illustration, consider the case of D&O insurance. Here, behaviors $b = 1, \dots, n$ correspond to all the possible ways through which corporate directors and officers may manage the firm they are in charge of, paying more or less attention to the legitimate interest of stockholders, customers and other stakeholders. A loss occurs (most often after a class action) when directors or managers are held liable by a court for compensating stakeholders for their loss. In this case, $\theta \in \Theta$ corresponds to all the relevant facts that have influenced the management and are unknown ex ante (i.e. occurring after the conclusion of the insurance contract). This refers to business situations that affect the behavior of directors and officers during the validity period of the insurance coverage.^{14,15}

¹³Assuming $d_b(\theta) > d_{b+1}(\theta)$ is not restrictive because the policyholder would prefer b to $b + 1$ in state θ if $d_b(\theta) \leq d_{b+1}(\theta)$ and $\pi_b < \pi_{b+1}$. In other words, behavior $b + 1$ would not be chosen in state θ .

¹⁴The agency approach to corporate finance emphasizes the endogeneity of management decisions, such as engaging in too risky projects, or giving undue preference to particular suppliers or subcontractors. Such misconducts may lead to lower expected profit and ultimately to financial distress, affecting shareholders and possibly employees of the firm. In addition to managers' moral standard, the possibility to earn such private benefits results from fluctuating business uncertainties, e.g., available free cash flows masking perks and privileges, or opportunities to contract with inefficient suppliers or subcontractors offering future job opportunities.

¹⁵Another illustration is given by the case of car insurance. A car driver may exert a low level of

We assume that θ is distributed in Θ according to a probability measure $H(\theta)$.¹⁶ 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 $u' > 0$ and $u'' < 0$. The insurer may collect evidence concerning 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, and $x = 0$ when there is no claim. Again, using D&O insurance as an example, this index would convey the informational value of the circumstances (such as undue personal profit or non-compliance with written injunctions) suggesting that the loss is likely to have resulted from a deliberate misbehavior of the insured, rather than from the intrinsic uncertainty of business life.

Index x is distributed in $[0, 1]$ according to c.d.f. $G_b(x)$ where $b \in \mathcal{B}$ is the policyholder's behavior, with $x = 0$ if there is no claim, and thus a mass of probability $G_b(0) = 1 - \pi_b$, and $x > 0$ in the case of an accident, with continuous derivative $g_b(x) = G'_b(x)$ if $x \in (0, 1]$. The p.d.f. and c.d.f. of x for behavior b conditionally on a loss occurring are denoted $\hat{g}_b(x) = g_b(x)/\pi_b$ and $\hat{G}_b(x) = \int_0^x \hat{g}_b(t)dt$, respectively. We assume that an increase in b induces a shift in the distribution of x , with strict MLRP when a loss occurs, i.e. $\hat{g}_{b+1}(x)/\hat{g}_b(x)$ is increasing with respect to x in $(0, 1]$ for all $b \in \{1, \dots, n-1\}$. In words, a large x leads one to think that the policyholder was at fault for severe misconduct.^{17,18} By an abuse of language, in what follows we may

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. All possible misconducts correspond to $b \in \{2, \dots, n\}$. 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 is not so bad to break the speed limit if he has to arrive on time at an important business meeting.

¹⁶We do not make any particular assumption on $H(\theta)$. We may assume, for instance, that it corresponds to a density $h(\theta)$. In Section 4, we consider the case where Θ includes a finite number of states.

¹⁷We 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 accident 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 $\varphi : \Omega \rightarrow [0, 1]$, such that x is a sufficient statistic for b and satisfies MLRP for all b . In particular, for any non-degenerate prior on b , an increase in x induces a FOSD shift in the posterior probability distribution of b . In this sense, a large x can be interpreted as a "bad news", i.e. as suggesting that the policyholder is guilty of serious misconduct.

¹⁸Since MLRP implies FOSD, we have $\hat{G}_{b+1}(x) < \hat{G}_b(x)$ for all x in $(0, 1)$. As $G_b(x) = 1 - \pi_b[1 - \hat{G}_b(x)]$ and $\pi_{b+1} > \pi_b$, we also have $G_{b+1}(x) < G_b(x)$ for all x in $[0, 1)$. To simplify matters (and to avoid corner solutions), we also assume $\hat{g}_b(x)/\hat{g}_1(x) \rightarrow 0$ when $x \rightarrow 0$ and $\hat{g}_b(x)/\hat{g}_1(x) \rightarrow \infty$ when $x \rightarrow 1$ if $b > 1$. Intuitively, $b = 1$ is much more likely than $b > 1$ when x is close to 0, and conversely when x is close to 1.

refer to x as the circumstances of the loss, although this is only the suspicion index derived from them. To summarize, 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 $b^*(\cdot) : \Theta \rightarrow \mathcal{B}$ the policyholder’s “behavioral rule”, i.e. $b^*(\theta)$ is the action taken by the insured in state $\theta \in \Theta$.¹⁹ 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. Under competitive forces, an equilibrium insurance contract maximizes the individual’s expected utility (i.e. her expected utility before she knows the state θ in which she will find herself) 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 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, they both observe the state θ and the circumstances x of any accident which may occur. This information (θ, x) is costlessly verifiable and may be used to condition the insurance coverage. Let $I(\theta, x)$ denote 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

¹⁹In particular, an insurance policy inducing a behavior $b^*(\theta) = 1$ for all θ is usually suboptimal because $d_1(\theta)$ may be very large in some states θ . In what follows, we just consider behavioral rules for which $b^*(\theta) = 1$ in a positive-measure subset of Θ .

with behavior $b \in \mathcal{B}$ is written as $u_b(\theta) - d_b(\theta)$, where

$$u_b(\theta) = (1 - \pi_b)u(W - P) + \pi_b \int_{0+}^1 u(W - P - L + I(\theta, x))\hat{g}_b(x)dx. \quad (1)$$

The first and second terms in (1) correspond to the no-accident and accident states respectively. If the insurance contract induces behavior $b^*(\theta)$ 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

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

and

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

respectively. The insurer's break-even constraint, and the no-overinsurance and non-negativity constraints are written as

$$P \geq \int_{\Theta} \pi_{b^*(\theta)} \int_{0+}^1 I(\theta, x)\hat{g}_{b^*(\theta)}(x)dx dH(\theta), \quad (4)$$

and

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

respectively.

The optimal insurance contract maximizes the expected utility given by (2) with respect to P , $I(\cdot)$ and $b^*(\cdot)$ subject to (3),(4) and (5).

Unsurprisingly, as shown in Proposition 1, $I^*(\theta, x)$ actually depends on θ and x . Indeed, for a given behavioral rule $b^*(\cdot)$, the uncertainty about θ should be taken into account in the insurance coverage. 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 $b^*(\theta) < n$, there exist $\underline{x}(\theta), \bar{x}(\theta) \in [0, 1]$ with $\underline{x}(\theta) < \bar{x}(\theta)$, such that the second-best optimal indemnity schedule $I^*(\theta, x)$ is continuous*

in x , with

$$\left. \begin{aligned} I^*(\theta, x) &= L && \text{if } 0 < x < \underline{x}(\theta) \ \&\ \underline{x}(\theta) > 0, \\ 0 < I^*(\theta, x) < L \\ dI^*(\theta, x)/dx < 0 \end{aligned} \right\} \text{if } \underline{x}(\theta) < x < \bar{x}(\theta),$$

$$I^*(\theta, x) = 0 \quad \text{if } \bar{x}(\theta) < x \leq 1 \ \&\ \bar{x}(\theta) < 1.$$

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

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

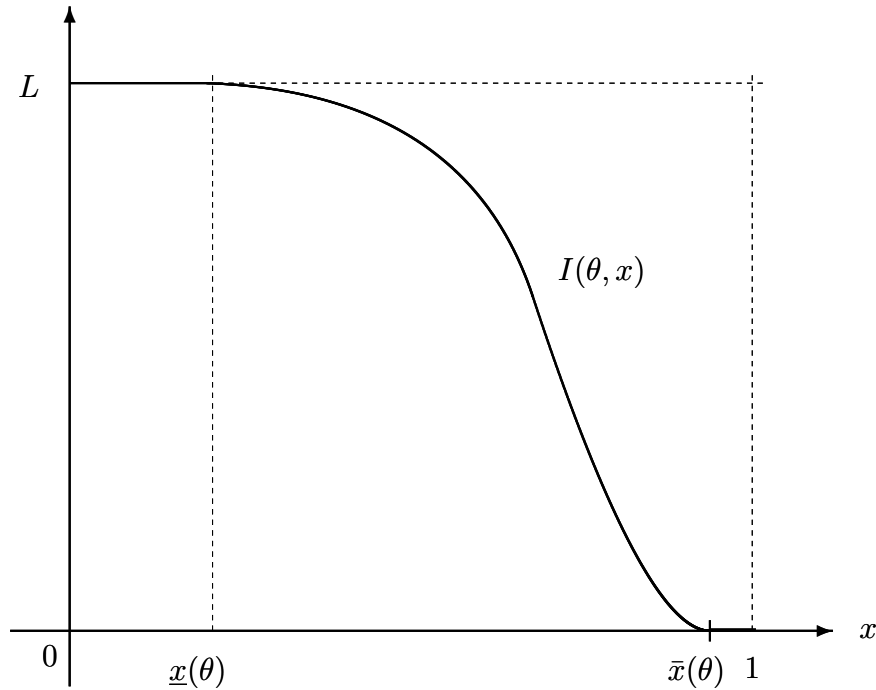


Figure 1: Optimal second-best indemnity

3 Incomplete contracts

Aside from 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 through exclusions for specific types of accident. Firstly, obtaining verifiable information about the circumstances of an accident usually requires a costly verification process. More often than not, claim handlers 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 surrounding it. Another reason is related to 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 legal principles governing insurance contracts, or on policyholders' duties.²⁰ 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.²¹

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 justify a more or less severe reduction in the indemnity. Indeed, even if competition between the insurance companies leads them to offer contracts that maximize the policyholder surplus ex ante, they are also tempted to behave opportunistically ex post, i.e. to deny coverage once the consumer is locked in. 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.

We consider in the following that the resolution of conflicts between the parties is left to the judge and that law-makers design the insurance law (i.e. the legal means through which the insurer may challenge the legitimacy of a claim) in order to maximize

²⁰For instance, calling for the service of affiliated car repairers or health care providers.

²¹Put differently, by using the terminology of cognitive science, complete contracts would induce such a large cognitive load (i.e. too much effort in using working memory) that their incentive advantages would not be worth it.

the social welfare.²² Put differently, we analyze the outcome of competitive forces in the insurance industry when the enforcement of incomplete insurance policies is governed by legal principles stated by benevolent law-makers. As the auditing cost coupled with the incompleteness of insurance policies are at the origin of an efficiency loss by comparison with the second-best solution, the resulting allocations correspond to the third-best solution to the insurance problem under moral hazard.

A standard way in the analysis of conflicts arbitrated by law consists in 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 to be bad faith on the part of the insurer and they would be invalidated. This consistency requirement corresponds to a standard of proof, such as the balance of probabilities, that is used for civil cases in Common Law.

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 density $f(\varepsilon)$ defined over \mathbb{R} and such that $\text{Cov}(x, \varepsilon) = 0$. 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, in which case we have $W_f = W - P - L + I$. If x has been verified through an audit, then the insurance payment depends on x through legal means that may be invoked by the insurer.²³ For notational simplicity we will not distinguish alleged misconduct $b \in \{2, \dots, n\}$ from the corresponding legal means (or broad standards) that can be invoked by insurers, although, as we have noted above, in practice there are many types of misconduct, while the law of insurance contracts only includes a limited number of legal means. Insurance law specifies the insurer's leeway in the claim settlement process, i.e. to what extent a legal means allows him to reduce or even cancel coverage. More

²²Similar results are obtained assuming state-sponsored or private conciliation institutions, as long as they can commit on a pre-announced standard of proof and on guiding principles that apply unambiguously to insureds and insurers.

²³In what follows, we assume that the policyholder is informed that her claim is audited. This is a plausible assumption since auditing usually requires additional information obtained from the claimant or witnesses, or other stakeholders.

precisely, in what follows, the law of insurance contracts is subsumed in the proportion $y_b \in [0, 1]$ of the contractual indemnity that the insurer is allowed to reduce 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 reduce 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 reductions are co-monotone, i.e. $y_b \leq y_{b+1}$ for all $b = 1, \dots, n-1$. If the insurer is in a position to invoke legal means b under circumstances x , then he may decide to reduce the indemnity by a fraction $z(x)$ lower or equal to y_b , and the insurance payment is $[1 - z(x)]I$, with $W_f = W - P - L + [1 - z(x)]I$.²⁴

When policyholders follow the behavior rule $b^*(\theta)$, 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

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

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

The standard of proof in use is characterized by parameter $\alpha \in [1/2, 1)$. Given x , misconduct $b_0 \in \{2, \dots, n\}$ is said to be “credibly alleged according to the standard of proof α ” if the probability that the policyholder had misconduct b_0 or a worse misconduct $b \in \{b_0 + 1, \dots, n\}$ is larger than α , i.e. if

$$\sum_{b=b_0}^n \Pr(b|x) > \alpha. \quad (7)$$

Allegation b_0 is said to be “credible on a balance of probabilities” when $\alpha = 1/2$, or equivalently when

$$\sum_{b=b_0}^n \Pr(b|x) > \sum_{b=1}^{b_0-1} \Pr(b|x),$$

In other words, under the balance of probabilities, it is more likely that the policyholder had misconduct b_0 or a worse misconduct $b \in \{b_0 + 1, \dots, n\}$ than a better behavior $b \in \{1, \dots, b_0 - 1\}$. In what follows, most of our results are robust to changes in the standard of proof, and parameter $\alpha \in [1/2, 1)$ is considered as given, without further specification.²⁵

²⁴Since our perspective is predominantly normative, we do not restrict the set of indemnity reductions that courts could enforce when misconduct b is credibly alleged. However, we will also consider the case of the so-called “negligence rule” where y_b is restricted to 0 or 1, according to whether the policyholder meets some standard of care, or not.

²⁵Case $\alpha = 1$ is excluded because it would mean that courts approve an allegation only when it can

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

$$\sum_{b=\hat{b}(x)+1}^n \Pr(b|x) \leq \alpha < \sum_{b=\hat{b}(x)}^n \Pr(b|x), \quad (8)$$

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

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

Lemma 1 allows us to write $\hat{b}(x) = b$ if $x_b < x \leq x_{b+1}$, $b = 1, \dots, m$ with $m \leq n$, $x_1 = 0$, $x_m + 1 = 1$, and x_b given by

$$\sum_{b'=b}^n \Pr(b'|x_b) = \alpha$$

for all $b = 2, \dots, m$, where m is the most credible severe misconduct under the standard of proof α , when x is close to 1. The corresponding maximum indemnity reductions are given by $y_{\hat{b}(x)}$ as illustrated in Figure 2 (where $m = 4$).

For example, when an audit reveals $x \in (x_3, x_4]$, the insurer is in a position to credibly allege a misconduct b_3 or worse, and law allows him to reduce the indemnity by a fraction y_3 without being challenged by a judge in the event of a court action by the insured. The claim may be fully denied when $x > x_4$.²⁶

The interactions between the judicial system (law-makers and judges), insurers and insureds can then be described by the following 6-stage game:²⁷

- **Stage 0:** Law-makers decide on $\{y_b, b = 2, \dots, n\}$.

be confirmed with full certainty. This would require that the support of x in $(0, 1]$ depends on behavior b , an unnecessarily restrictive assumption. Proof “beyond reasonable doubt” would corresponds to α close to 1.

²⁶The thresholds x_2, x_3, \dots, x_m depend on the standard of proof α and are associated with the behavioral rule $b^*(\cdot)$.

²⁷For the sake of clarity, we distinguish law-makers who decide in which cases and to what extent insurers are allowed to deny coverage, from judges who apply these law principles under a standard of proof. Most importantly, we assume that law-makers are acting as a social planner: they choose the law principles in order to maximize the ex ante policyholders’ expected utility. Note that we consider the standard of proof α as exogenously given, since, in practice, the same standard of proof is applied to a wide class of law suits, like the balance of probabilities for civil cases under Common Law.

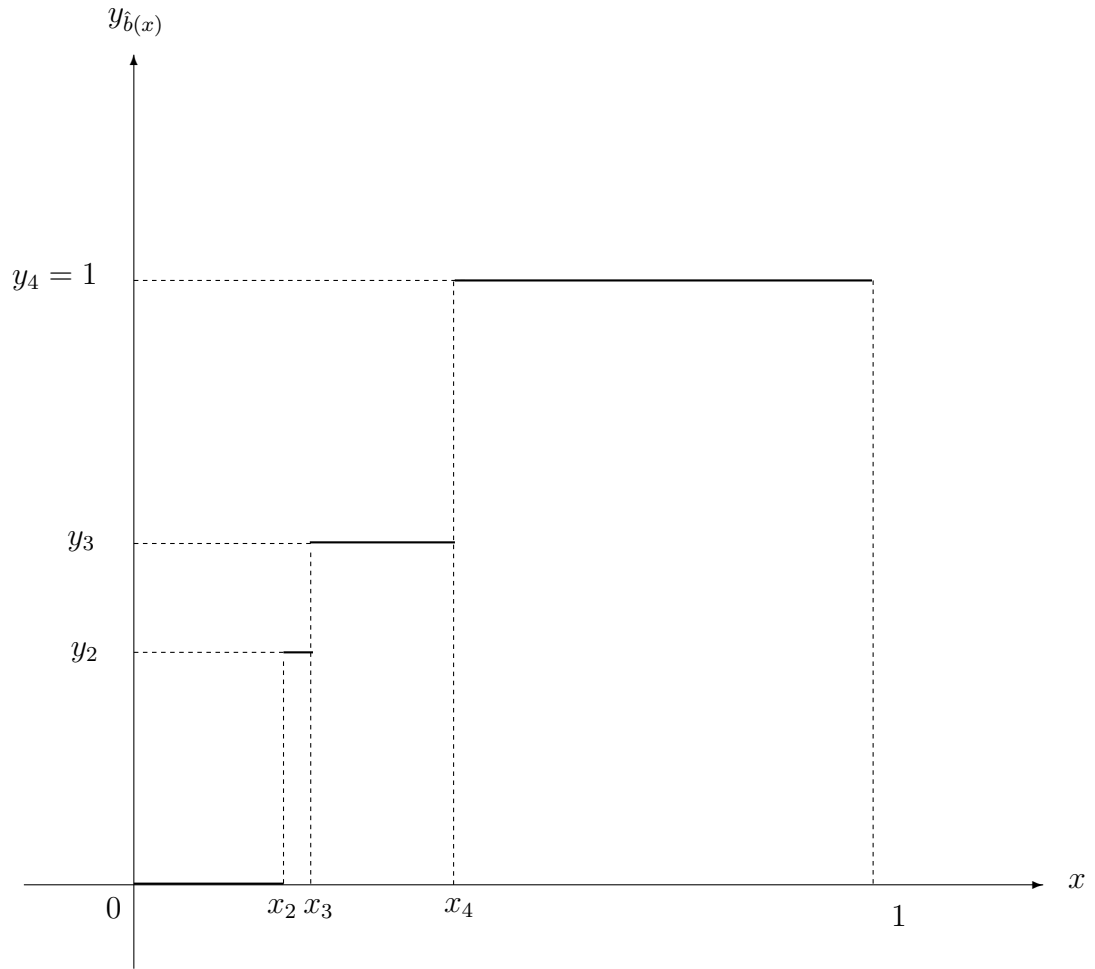


Figure 2: Signal thresholds and indemnity reductions.

- **Stage 1:** Each insurer offers an insurance contract (I, P) .
- **Stage 2:** Each individual takes out her preferred insurance policy among the offers, and nature choose θ for each policyholder. The policyholder observes θ and then chooses behavior $b \in \mathcal{B}$. Should a loss occur, she observes its circumstances x and files a claim. In that case, the insurer observes signal s .
- **Stage 3:** If a claim has been filed, the insurer either pays I or triggers an audit. In that case, he incurs the audit cost c and he obtains the verifiable information x . The policyholder observes whether her claim is audited or not.
- **Stage 4:** If an audit has been performed, the insurer either pays I to the claimant, or he alleges that the policyholder misbehaved according to $b \in \{2, \dots, n\}$.
- **Stage 5:** The policyholder may contest the insurer's allegation in court. The judge

confirms the insurer's allegation if $b \leq \hat{b}(x)$, and dismisses it otherwise.

- **Stage 6:** The indemnity paid to the policyholder is I 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 2, the insurer triggers an audit at stage 3 if $c \leq \mathbb{E}[y_{\hat{b}(x)}|s]I$, since $b = \hat{b}(x)$ is the most severe allegation made at stage 4 that will not be dismissed by the judge at stage 5, and the insurer chooses $z = y_{\hat{b}(x)}$ at stage 6. Thus, an equilibrium audit strategy is defined by

$$q(s) = \begin{cases} 1 & \text{if } c \leq \mathbb{E}[y_{\hat{b}(x)}|s]I, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

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

The intuition of lemma 2 is straightforward. Its proof goes through preliminary Lemma 3 in the Appendix C, showing that under log-concavity of $f(\varepsilon)$ an increase in s induces a FOSD shift in the conditional probability distribution of x . It is worth auditing a claim when the signal s is bad enough to be considered to be a red flag, indicating that the circumstances of the loss are likely to be unfavorable (i.e. x is probably large). Cases where $s^* = \pm\infty$ correspond to corner solutions where claims are never (resp. always) audited because c is very large (resp. very low). The rest of the analysis is performed assuming that $f(\varepsilon)$ is log-concave.²⁸

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

$$\begin{aligned} u_b &= (1 - \pi_b)u(W - P) + \pi_b \int_{0_+}^1 u(W - P - L + I) (1 - \mathbb{E}[q(s)|x]) \hat{g}_b(x) dx \\ &+ \pi_b \int_{0_+}^1 u(W - P - L + (1 - z(x))I) \mathbb{E}[q(s)|x] \hat{g}_b(x) dx \end{aligned} \quad (10)$$

with $\mathbb{E}[q(s)|x] = \int_{-\infty}^{+\infty} q(s) f(s - x) ds$ and

$$z(x) = y_{\hat{b}(x)} \quad (11)$$

²⁸Log-concave densities are unimodal, with convex level sets and sub-exponential tails. Many univariate random variables whose distributions are defined by densities are log-concave, including the Normal, Laplace, Uniform, Gamma and Beta distributions.

for all x . The first, second and third terms in (10) correspond to the no-accident state, to the accident states without audit and to the accident states with audit, respectively.

Since law-makers are supposed to be benevolent (they aim at maximizing the ex ante expected utility of policyholders) and identical insurers interact in a competitive market, the equilibrium values of insurance contract (P, I) , audit strategy $q(\cdot)$, and insurance law $\{y_b, b \in \mathcal{B}\}$ jointly maximize

$$\mathbb{E}u^* \equiv \int_{\Theta} [u_{b^*(\theta)} - d_{b^*(\theta)}(\theta)] dH(\theta) \quad (12)$$

subject to

$$P \geq \int_{\Theta} \pi_{b^*(\theta)} \left\{ I + \int_{0+}^1 (c - Iy_{\hat{b}(x)}) \mathbb{E}[q(s)|x] \hat{g}_{b^*(\theta)}(x) dx \right\} dH(\theta), \quad (13)$$

$$u_{b^*(\theta)} - d_{b^*(\theta)}(\theta) \geq u_b - d_b(\theta) \text{ for all } (\theta, b) \in \Theta \times \mathcal{B}, \quad (14)$$

$$q(s) = \begin{cases} 1 & \text{if } s \geq s^* \\ 0 & \text{if } s < s^* \end{cases}, \quad (15)$$

where s^* is given by

$$c = \mathbb{E}[y_{\hat{b}(x)} | s^*] I, \quad (16)$$

and $\hat{b}(\cdot) : [0, 1] \rightarrow \mathcal{B}$, $b^*(\cdot) : \Theta \rightarrow \mathcal{B}$ satisfy conditions (6) and (8).

Notations can be recapped as follows: $b^*(\theta)$ is the policyholder's behavior in state θ , $\hat{b}(x)$ is the most severe policyholder's misconduct that can be credibly alleged by the insurer after observing circumstances x through a claim audit, and s^* is the signal threshold above which an audit is triggered. Hence, $\mathbb{E}u^*$ given by (12) is the ex ante policyholder's expected utility, i.e. before she knows the state θ , with behavioral rule $b^*(\cdot)$. Condition (15) is deduced from condition (9) and Lemma 2: an audit is triggered if $s \geq s^*$, where s^* is given by (16).²⁹ 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, (13) is the insurer's break-even condition. (14) is the ex post incentive constraint: in state θ , the policyholder weakly prefers behavior $b^*(\theta)$, rather than any other behavior $b \neq b^*(\theta)$.^{30,31}

²⁹Eq. (16) assumes that $s^* \neq \pm\infty$. It can be easily extended to these cases.

³⁰In what follows, in order to avoid corner solutions where the insured would always choose the best or worst behavior, we assume that the no-audit optimal strategy $b_{na}^*(\theta)$ —i.e. the optimal solution under $z(x) = 0$ for all x — is such that $1 < b_{na}^*(\theta) < n$ in a positive-measure subset of Θ . We also assume that $I > 0$ at such a no-audit solution.

³¹At a third-best optimal solution, policyholders may not choose their most efficient behavior (i.e.

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. $s^* < +\infty$).*

In short, when court decisions are based on a standard of proof $\alpha \in [1/2, 1)$, auditing claims puts the insurer in the position to credibly allege that the policyholder misbehaved in some way when x is sufficiently large, i.e. when $\hat{b}(x) \geq 2$. The insurer will not be dismissed by the court and, in such cases, insurance law should allow him to reduce or cancel indemnities. It is worth auditing when the expected claim reduction $E[y_{\hat{b}(x)}|s]I$ is larger than the audit cost c , which holds when $s \geq s^*$, with $s^* < +\infty$ if c is not too large. This leeway provided by 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, although it is not written in contractual clauses. From this standpoint, the variability of coverage should not be viewed as reflecting the opportunistic behavior of insurers (as often suggested by the bad reputations of nitpicky insurers), but as the outcome of law provisions improving the efficiency of insurance mechanisms under moral hazard when contracts are incomplete.

Proposition 2 is established without preconditions on the proportions y_2, \dots, y_n , apart from the fact that they are non-decreasing. Additional constraints may be imposed by more specific rules of law. This is the case of the “negligence rule” which play an important role in liability law, when acceptable standards of safe conduct can be defined and applied to a variety of settings. In our framework, the negligence rule is defined by a “standard of care” \bar{b} : the insurer is allowed to fully cancel the coverage if he can credibly allege (under the current standard of proof) that the insured misbehaved by choosing $b > \bar{b}$, or equivalently that her accident probability was larger than

the lowest accident probability) among the behaviors that are equivalent to them. The reason for this paradox is that the behavior rule $b^*(\cdot)$ affects $\hat{b}(\cdot)$, and thus function $z(\cdot)$. The proofs show that Propositions 2 and 3 remain valid if we postulate that policyholders always choose their lowest accident probability among their optimal choices. More generally, the model could be extended to the case where policyholders randomize among their best choices. This would reinforce the conclusions of Propositions 2 and 3, because such mixed strategies are not optimal in the standard moral hazard model without auditing. In an insightful remark, a referee observed that allowing for mixed-strategy could be a way to extend our results to the case of a single type (i.e. Θ is a singleton). However, assuming that players coordinate themselves to choose pure strategies with equilibrium probabilities (each pure strategy being chosen by a fraction of agents), is a problem in its own right. Our approach is in the spirit of the correlated equilibrium introduced by Aumann (1974, 1987) since the distribution of pure equilibrium strategies results from signal θ acting as a strategy generator.

$\pi_{\bar{b}}$, and he has to pay the full indemnity otherwise. Hence, the negligence rule gives

$$\begin{aligned} y_b &= 1 \text{ if } b > \bar{b}, \\ y_b &= 0 \text{ if } b \leq \bar{b}. \end{aligned}$$

It can be enforced by denying coverage under circumstances $x > x_{\bar{b}+1}$, where x_1, x_2, \dots, x_m denote the discontinuity points of function $\hat{b}(x)$. In other words, $x_{\bar{b}+1}$ is the threshold above which it can be credibly alleged that the insured's behavior was $\bar{b} + 1$ or worse. This gives $z(x) = 0$ if $x \leq x_{\bar{b}+1}$ and $z(x) = 1$ if $x > x_{\bar{b}+1}$. In short, a negligence rule restricts the stipulations of law to an all-or-nothing alternative, the enforcement of the standard of care \bar{b} being constrained by the credibility constraint associated with the standard of proof α .

Not surprisingly, this restriction weakens the effectiveness of legal mechanisms under moral hazard, hence leading to a lower policyholder's expected utility, for α unchanged. The following proposition provides sufficient conditions for the negligence rule to be nevertheless more efficient than the no-audit solution.

Proposition 3 *Assume $\lim_{x \rightarrow 1} g_{b+1}(x)/g_b(x) \rightarrow +\infty$ for all $b \in \{1, \dots, n-1\}$ and c is not prohibitively large. There exists a standard of care $\bar{b} < n$ and a standard of proof $\bar{\alpha} \in [1/2, 1)$ such that, when $\alpha \geq \bar{\alpha}$, an optimal third-best solution to the insurance moral hazard problem under the negligence rule dominates the no-audit optimal solution, with positive probability of audit (i.e. $s^* < +\infty$).*

The first assumption in Proposition 3 conveys the idea that the worst signals (i.e. x close to 1) are very informative about the severity of misconducts. Intuitively, when x tends to 1, any misconduct $b + 1$ becomes much more likely than b . Hence, starting from the no-audit allocation, intuition suggests that an efficiency gain could be reached by a negligence rule with standard of care $\bar{b} < m_{na}(\alpha)$. Here, $m_{na}(\alpha)$ is the most severe misconduct in the optimal non-audit solution that can be credibly alleged under the standard of proof α . However, in some cases, matters may be less simple than expected. Consider the case where an important proportion of policyholders choose $b = m_{na}(\alpha)$ at the optimal no-audit solution. In that case, $x_{m_{na}(\alpha)}$ may be relatively small, i.e. for many values of x , the insurer can credibly allege that the policyholder's behavior was $b = m_{na}(\alpha)$ or worse. Hence, adopting the negligence rule $\bar{b} = m_{na}(\alpha) - 1$ may be suboptimal because it would indifferently affect a large proportion of policyholders. This argument is valid for a given standard of proof $\alpha \in [1/2, 1)$, and in particular for the balance of probabilities criterion $\alpha = 1/2$. Increasing the standard of proof

α is a way to smooth the effect of the negligence rule when its effects are too harsh. When α is raised, the thresholds x_b increase and the possible negative effects of the negligence rule are mitigated. Under this caveat of a sufficiently stringent standard of proof, Proposition 3 shows that the negligence rule also induces an efficiency gain by comparison with the no-audit solution, provided that signals are very informative about the worst misconducts.

4 Illustrative examples

In this section, we assume a discrete distribution of states., i.e. $\Theta = \{\theta_1, \dots, \theta_n\}$ with probabilities $h_i = \Pr\{\theta = \theta_i\} > 0$, $\sum_{i=1}^n h_i = 1$, and $n = 2$ or 3 . We characterize the optimal third-best insurance contract corresponding to separating allocations where the policyholder chooses b_i in state θ_i . We assume that the audit cost is negligible, i.e. $c = 0$, but conclusions remain valid if c is not too large.

4.1 The two-state case

Assume $\mathcal{B} = \{1, 2\}$ and $\Theta = \{\theta_1, \theta_2\}$ with $d_1(\theta_1) > d_2(\theta_1) > 0$ and $d_1(\theta_2) = +\infty, d_2(\theta_2) \geq 0$. In words, θ_1 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 θ_2 . We consider an optimal third-best allocation where the policyholder chooses $b = 1$ when $\theta = \theta_1$, and $b = 2$ when $\theta = \theta_2$. 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 characterizes separating allocations and is deduced from $\Pr(b = 2|x_2) = \alpha$, i.e.

$$\frac{\hat{g}_2(x_2)}{\hat{g}_1(x_2)} = \frac{\alpha \pi_1 h_1}{1 - \alpha \pi_2 h_2}. \quad (17)$$

Eq. (17) defines implicitly a threshold function $x_2(\alpha)$ that increases with α from $x_2(0) = 0$ to $x_2(1) = 1$. It is convenient to restate our problem in terms of utils rather than monetary values. This is done by defining $\hat{u} = u(W - P)$, $u_0 = u(W - P - L + I)$, $u_1 = u(W - P - L + (1 - y_2)I)$, and $v = u_0 - u_1$. Since the type- θ_1 incentive constraint is binding, it may be rewritten as

$$\hat{u} = u_0 + A - v\Delta \quad (18)$$

where $\Delta = [G_1(x_2) - G_2(x_2)]/(\pi_2 - \pi_1) > 0$ and $A = [d_1(\theta_1) - d_2(\theta_1)]/(\pi_2 - \pi_1) > 0$. The insurer's break-even constraint can be expressed as

$$\begin{aligned} \bar{W} &\geq u^{-1}(u_0) \sum_{i=1,2} h_i \pi_i \hat{G}_i(x_2) + u^{-1}(u_0 - v) \sum_{i=1,2} h_i \pi_i (1 - \hat{G}_i(x_2)) \\ &\quad + \left(1 - \sum_{i=1,2} h_i \pi_i\right) u^{-1}(u_0 + A - v\Delta) \\ &\equiv K_1(u_0, v), \end{aligned} \tag{19}$$

where $\bar{W} = W - L \sum_{i=1,2} h_i \pi_i$ is the ex ante policyholder's expected wealth. Since u is concave, u^{-1} is convex and thus $K_1(u_0, v)$ is a convex function of u_0 and v , with $\partial K_1/\partial u_0 > 0$ and $\partial K_1/\partial v < 0$. The same substitutions yield

$$\mathbb{E}u^* = \sum_{i=1,2} h_i \{(1 - \pi_i)(u_0 + A - v\Delta) + \pi_i[u_0 - (1 - \hat{G}_i(x_2))v] - d_i(\theta_i)\} \tag{20}$$

using (18). Finally, the conditions $y_2 \geq 0$ and $I \leq L$ can be written as $v \geq 0$ and $u_0 \leq \hat{u}$, respectively. We thus must have³²

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

Given a standard of proof α , the optimal solution is obtained by maximizing $\mathbb{E}u^*$ given by (20) w.r.t. u_0 and v , subject to (19) and (21). As illustrated in Figures 3 and 4, the insurer's break-even constraint $K_1(u_0, v) = \bar{W}$ corresponds to an increasing convex locus in the (u_0, v) plane. Iso-expected utility curves are increasing straight lines. Point D_1 corresponds to $v = 0$, hence $y_2 = 0$: the insurer is not allowed to cut the indemnity, and the optimal solution to the moral hazard problem only goes through the deductible $L - I$. On the contrary, E_1 corresponds to $I = L$ and $y_2 > 0$. When moving from D_1 to E_1 on the insurer's break-even locus, the legal penalty v is progressively substituted to the deductible $L - I$, and the latter fully vanishes at point E_1 . Figure 3 illustrates the case of a corner solution at point E_1 where $v = A/\Delta$, and thus with $I = L$. For a given standard of proof α , it is shown in the appendix that this is the case when h_2 is large enough. More explicitly:

Proposition 4 *Assume $n = 2$. We have $y_2 > 0$. Furthermore, there exists $h^* \in (0, 1)$, such that the optimal third-best contract provides full coverage (i.e. $I = L$) if $h_2 \geq h^*$ and partial coverage (i.e. $I < L$) if $h_2 < h^*$.*

³²We must also have $y_2 \leq 1$. Appendix F shows that this is the case if A is small enough.

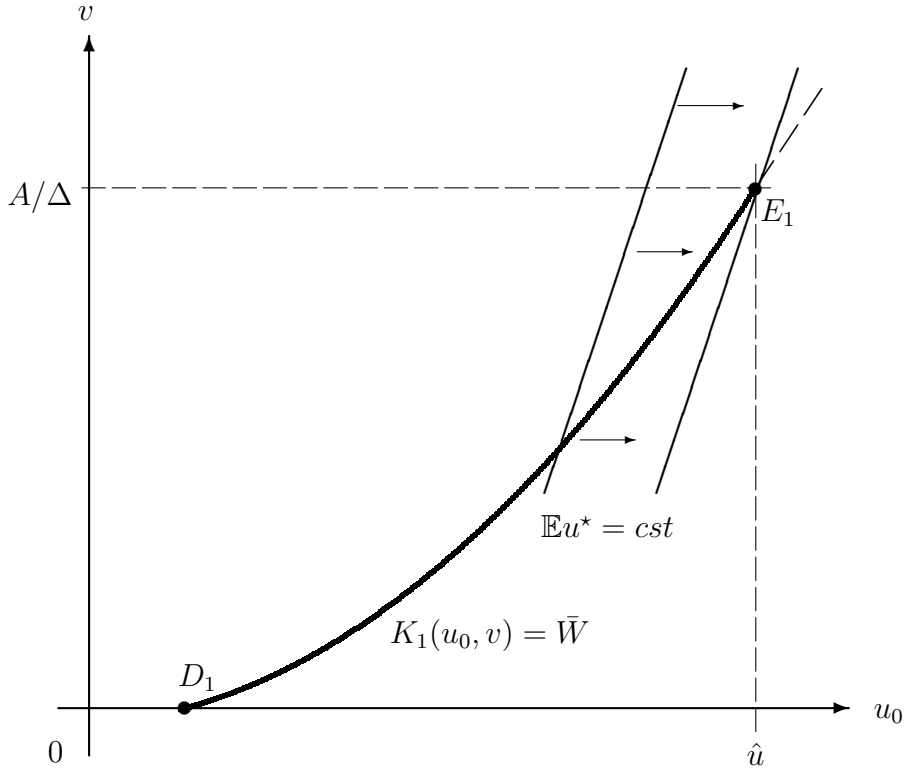


Figure 3: Equilibrium contract when $h_2 > h^*$.

The first part of Proposition 4 just rephrases Proposition 2 when $n = 2$, but with a simple proof provided in appendix H. The following of Proposition 4 highlights the trade-off between the incentives provided by the law of insurance contracts and those provided by the indemnity schedule. Two regimes exist, according to whether h_2 is larger or smaller than the threshold h^* .³³ If many individuals exert the low effort level (i.e. if $h_2 > h^*$), it is optimal to offer an insurance contract with full coverage and to allow the insurer to reduce or cancel the indemnity when the circumstances reveal misconduct.³⁴ On the contrary, if $h_2 < h^*$, it is difficult for the insurer to credibly claim that the insured did misbehave (i.e. this is possible only for very large x), and incentives are better provided by offering a contractual indemnity that is lower than the loss, whatever the circumstances. These two cases are illustrated by Figures 3 and

³³ h^* depends on $G_1(\cdot)$ and $G_2(\cdot)$, i.e. on the probability distributions of the circumstances of a loss.

³⁴The corner solution $I = L$ illustrated in Figure 3 shows that $I > L$ might be optimal if I were not bounded to be in $[0, L]$. Intuitively, an increase in I discourages the policyholder to exert effort, but it also increases the penalty $y_2 I$ when $x > x_2$, which incentivizes her to make more effort. This makes an optimal solution with $I > L$ possible.

4, respectively.

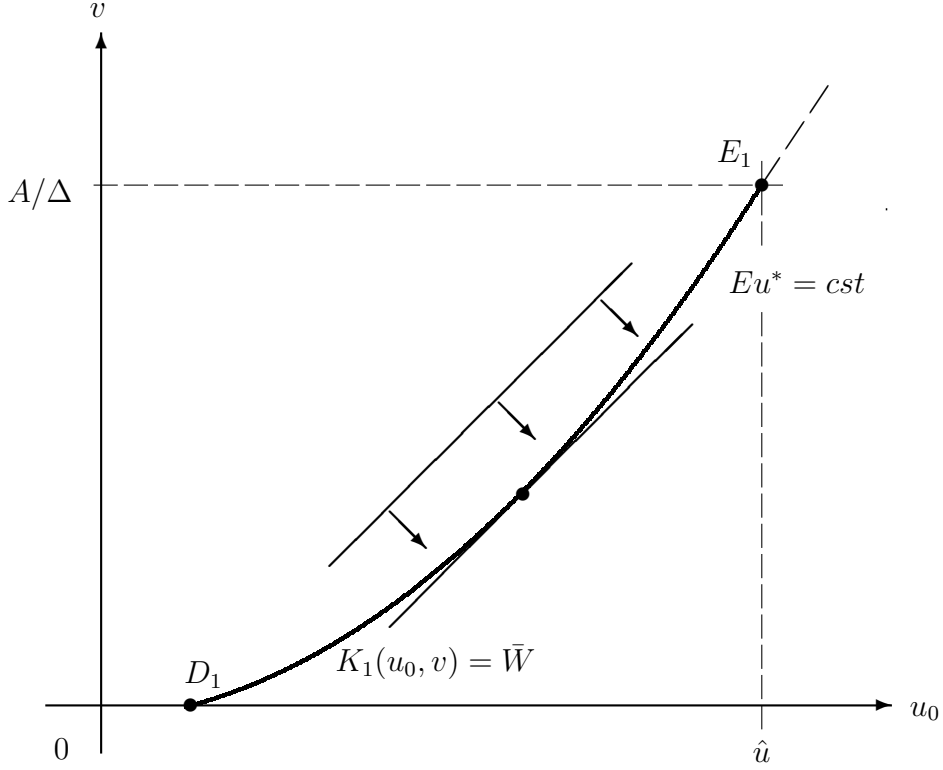


Figure 4: Equilibrium contract when $h_2 < h^*$.

The case of the negligence rule corresponds to the reduction function $z(x)$ restricted to either 0 or 1, hence, a reduced indemnity $(1 - y_2)I = 0$ and a corresponding utility level $u_1 = u(W - P - L)$. Similarly to (19), the insurer break-even constraint can be written as $\bar{W} \geq K_2(u_0, v)$ with

$$\begin{aligned}
 K_2(u_0, v) = & \left(1 - \sum_{i=1,2} h_i \pi_i \hat{G}_i(x_2) \right) u^{-1}(u_0 + A - v\Delta) \\
 & + u^{-1}(u_0) \sum_{i=1,2} h_i \pi_i \hat{G}_i(x_2) - L \sum_{i=1,2} h_i \pi_i [1 - \hat{G}_i(x_2)]
 \end{aligned} \tag{22}$$

which is a convex function of u_0 and v , satisfying $\partial K_2 \partial u_0 > 0$ and $\partial K_2 \partial v < 0$. We

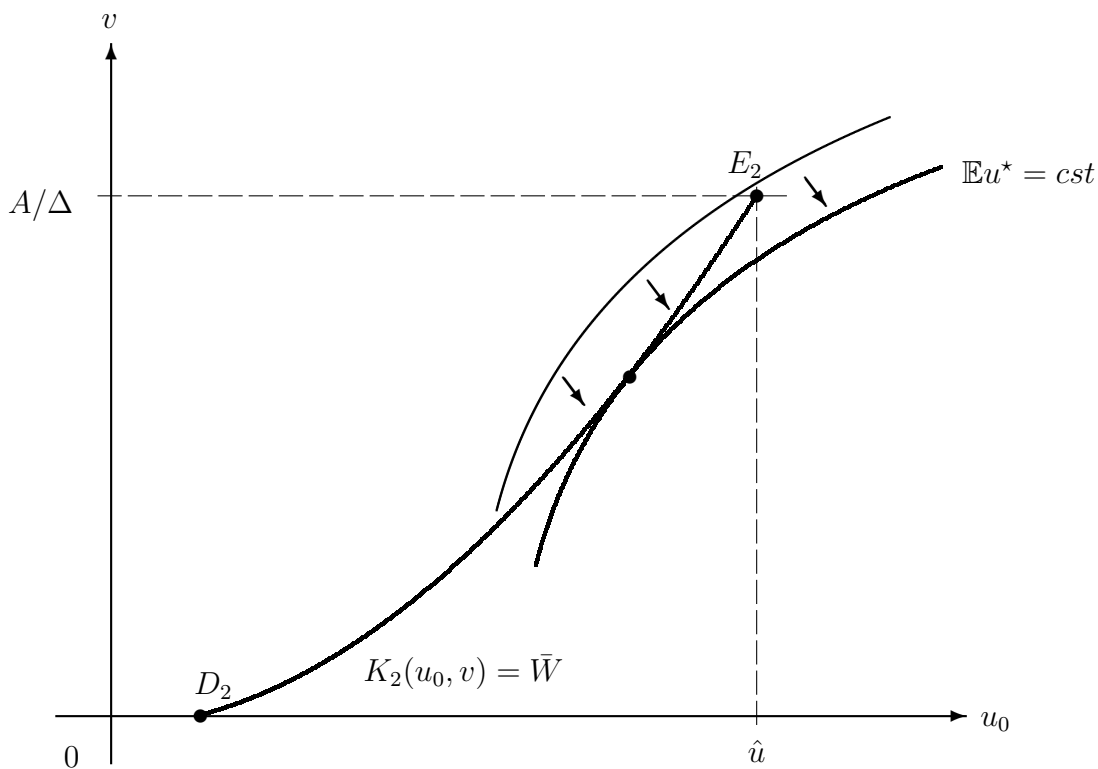


Figure 5: Equilibrium contract under the negligence rule.

also have

$$\begin{aligned}
\mathbb{E}u^* &= \sum_{i=1,2} h_i \{ (1 - \pi_i)(A - \Delta v + u_0) \\
&\quad + \pi_i [\hat{G}_i(x_2)u_0 + (1 - \hat{G}_i(x_2))u(u^{-1}(A - \Delta v + u_0) - L)] - d_i(\theta_i) \} \\
&\equiv U(u_0, v)
\end{aligned}$$

which is a concave function of u_0 and v if u is DARA, with $\partial U/\partial u_0 > 0$ and $\partial U/\partial v < 0$. Under the conditions given in Proposition 3, it is shown in the appendix that the negligence rule leads to $v < A/\Delta$, i.e. $I < L$, at the optimum, as illustrated in Figure 5.

Proposition 5 *Assume $n = 2$. There exists $h^{**} \in (0, 1)$ such that the optimal third-best solution under the negligence rule with $\bar{b} = 1$ dominates the no-audit optimal solution if $h_2 < h^{**}$. The corresponding optimal contract provides partial coverage, i.e. $I < L$.*

Proposition 5 states that the desirability of the negligence rule depends on h_2 being larger or smaller than threshold h^{**} . To intuitively interpret this result, observe that $\hat{g}_2(x)/\hat{g}_1(x) \rightarrow +\infty$ when $x \rightarrow 1$ yields $x_2 \rightarrow 1$ when $h_2 = 1 - h_1 \rightarrow 0$, everything else being equal. Thus, when h_2 is small, the negligence rule resembles an almost perfectly targeted penalty imposed only in case of misbehavior $b = 2$, hence the dominance result.³⁵ However, the all-or-nothing intrinsic feature of the negligence rule limits the scope of its incentive power, and it should be completed by partial coverage.

4.2 The three-state case

Suppose now that $\mathcal{B} = \{1, 2, 3\}$ and $\Theta = \{\theta_1, \theta_2, \theta_3\}$, hence two types of misconduct and three states, with $d_b(\theta_1) < d_b(\theta_2) < d_b(\theta_3)$ for all $b \in \mathcal{B}$. We assume that, at an optimal third-best allocation, the policyholder chooses $b^*(\theta_i) = b_i$, the adjacent ex post incentive constraints of types $i = 1, 2$ being binding (i.e. in state θ_i , the policyholder is incentivized not too choose b_{i+1}). Thresholds x_2 and x_3 satisfy $1 - \Pr(b = b_1|x = x_2) = \Pr(b = b_3|x = x_3) = \alpha$. We have

$$u_b = (1 - \pi_b)\hat{u} + \pi_b u_0 - \underline{v}\pi_b[1 - \hat{G}_b(x_2)] - \bar{v}\pi_b[1 - \hat{G}_b(x_3)],$$

where $\hat{u} = u(W - P)$, $u_0 = u(W - P - L + I)$, $\underline{v} = u_0 - u(W - P - L + (1 - y_2)I)$ and $\bar{v} = u(W - P - L + (1 - y_2)I) - u(W - P - L + (1 - y_3)I)$. We thus have $\underline{v} > 0$ and $\bar{v} > 0$ if $y_2 > 0$ and $y_3 > y_2$, respectively. Adjacent incentive constraints $u_2 - u_3 \geq d_2(\theta_2) - d_3(\theta_2)$ and $u_1 - u_2 \geq d_1(\theta_1) - d_2(\theta_1)$ may be written as

$$\hat{u} \geq \max_{i=1,2} \{u_0 + A_i - \underline{v}\Delta_i(x_2) - \bar{v}\Delta_i(x_3)\},$$

where $A_i = [d_i(\theta_i) - d_{i+1}(\theta_i)]/(\pi_{i+1} - \pi_i) > 0$ and $\Delta_i(x) = [G_i(x) - G_{i+1}(x)]/(\pi_{i+1} - \pi_i) > 0$. When both incentive constraints are binding, we have

$$M \cdot \begin{bmatrix} \underline{v} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} A_1 - z \\ A_2 - z \end{bmatrix}, \quad (23)$$

³⁵This is true under any standard of proof $\alpha \in [1/2, 1)$. To make a link with Proposition 3, note that for h_2 given, a large enough value of α also guarantees that x_2 is close to 1, with the same dominance conclusion.

where $z \equiv \hat{u} - u_0 \geq 0$, $z \leq \min\{A_1, A_2\}$, and

$$M = \begin{bmatrix} \Delta_1(x_2) & \Delta_1(x_3) \\ \Delta_2(x_2) & \Delta_2(x_3) \end{bmatrix}.$$

Appendix H characterizes a family of probability distributions that satisfies MLRP and such that $\det M = \Delta_1(x_2)\Delta_2(x_3) - \Delta_1(x_3)\Delta_2(x_2) > 0$. In Proposition 6 below, we assume that the probability distributions of x , conditional on $b = 1$ or 2 , belong to this family. It can be easily shown that $0 \leq z \leq Lu'(w - L) \equiv \hat{z}$, which leads to Proposition 6 that provides a sufficient condition under which $\underline{v} > 0$ and $\bar{v} > 0$ at a solution of (23).

Proposition 6 *Assume $n = 3$. If*

$$\frac{\Delta_2(x_3)}{\Delta_1(x_3)} > \frac{A_2}{A_1} > \frac{\Delta_2(x_2)}{\Delta_1(x_2)}, \quad (24)$$

then for \hat{z} small enough, we have $y_3 > y_2 > 0$ at any third-best optimal allocation where the adjacent incentive constraints are binding.

Proposition 6 provides sufficient conditions under which the optimal insurance law includes two levels of indemnity reductions, according to the severity of the misconducts that can be credibly alleged. An intuitive interpretation of this proposition is as follows. Let $\beta_i(x) \equiv \pi_{i+1}[1 - \hat{G}_{i+1}(x)] - \pi_i[1 - \hat{G}_i(x)]$ be the increase in the probability of a signal larger than $x > 0$ when the policyholder changes her behavior from $b = i$ to $b = i + 1$. In state i , $\beta_i(x_2)$ and $\beta_i(x_3)$ can be considered as indexes of the incentive power of the signal x evaluated at the thresholds x_2 and x_3 , respectively, these thresholds depending on α and the probabilities of states h_1, h_2 and h_3 . A simple calculation gives

$$\Delta'_i(x)/\Delta_i(x) = \beta'_i(x)/\beta_i(x). \quad (25)$$

The class of probability distributions characterized in Appendix H is such that

$$\beta'_1(x)/\beta_1(x) < \beta'_2(x)/\beta_2(x) < 0 \text{ for all } x, \quad (26)$$

Thus, for this class of probability distributions, when x is increasing, the incentive power index $\beta_i(x)$ decreases more strongly in state 1 than state 2. Intuitively, the low threshold x_2 then affects relatively more the policyholder in state 1 than in state 2, and conversely for the high threshold x_3 . Decreasing the indemnity in proportions

$y_2 > 0$ and $y_3 - y_2 > 0$ at the two successive thresholds x_2 and x_3 is a way to fine-tune the incentives in our multi-state setting. (25) and (26) yield $\Delta_2(x_2)/\Delta_1(x_2) < \Delta_2(x_3)/\Delta_1(x_3)$, hence the outer inequality in (24). The other inequalities depend on the costs of effort and on the loss, reflected by A_1, A_2 and \hat{z} , respectively. If they are satisfied, the separation of states through this fine-tuning of incentives is actually optimal.

5 The burden of proof

5.1 Litigation process and the cost of information

We consider in the following that providing evidence to the court is costly for both the insured and the insurer. For the latter, auditing a claim at cost c only provides soft information about the circumstances of the loss: 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).³⁶ Similarly, while she knows the circumstances x of the loss, sustaining a claim by verifiable information is costly to the policyholder, although presumably at a cost smaller than the one incurred by the insurer. Whatever the standard of proof α (considered as given in what follows), 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 and 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 the insured k_P per dollar of indemnity to produce evidence on the circumstances of the loss. We also assume that evidence can only be found about the actual events that generated the loss, i.e. that the information cannot be distorted or forged.³⁷ 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 insurance law design problem is modified as follows:

- **Stages 0 to 4** are unaffected, but the information x obtained by the insurer in case

³⁶For instance, 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.

³⁷Crocker and Morgan (1997) develop a theory of claim falsification, in which policyholders spend resources to distort the information insurers perceive about their loss.

of an audit is only soft information. Hence, the insurer's allegation $b \in \{2, \dots, n\}$ is not sustained by evidence, i.e. it does not allow the judge to verify the true value x .

- **Stage 5:** the policyholder may 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 the insurer allegation is not credible, i.e. $b > \hat{b}(x)$, 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 6** is unaffected.

In sum, the only difference with the game described Section 3 is that the policyholder incurs cost k_P to contest the insurer's allegation. 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 $\hat{b}(x) < b$, the judge considers that the insurer was plainly deceptive and he obliges him to pay the contractual indemnity as bad faith penalty.³⁸

After an audit that revealed x , the insurer knows that any allegation $b > \hat{b}(x)$ at stage 4 will be successfully contested by the insured at stage 5 if and only if $k_P \leq y_b$, because otherwise contesting the insurer's allegation would not be worthwhile. Let

$$x^* = \inf\{x \in (0, 1] : y_{\hat{b}(x)} \geq k_P\}. \quad (27)$$

When $x \geq x^*$, it is optimal for the policyholder to contest any allegation $b > \hat{b}(x)$, with net gain

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

Consequently, when the insurer observes $x \geq x^*$ 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 $x < x^*$, 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 $b \geq \hat{b}(x^*)$. Hence, an optimal insurer's strategy consists in alleging the misconduct immediately lower than $\hat{b}(x^*)$, i.e. $b = \hat{b}(x^*) - 1$, which will not be contested by the policyholder.

³⁸Should the insurer have alleged $b = \hat{b}(x)$, then he would have been allowed to reduce the indemnity by a fraction $y_{\hat{b}(x)}$.

The indemnity reduction is

$$z(x) = \begin{cases} y_{\hat{b}(x^*)-1} & \text{if } x < x^* \\ y_{\hat{b}(x)} & \text{if } x \geq x^* \end{cases} \quad (28)$$

instead of (11), and, whatever the case, the insured does not contest the insurer's allegation. The corresponding break-even condition for the insurer is given by

$$P \geq \int_{\Theta} \pi_{b^*(\theta)} \left\{ I + \int_{0_+}^1 cE[q(s)|x]\hat{g}_{b^*(\theta)}(x)dx - I \left(\int_{0_+}^{x^*} y_{\hat{b}(x^*)-1} + \int_{x^*}^1 y_{\hat{b}(x)} \right) E[q(s)|x]\hat{g}_{b^*(\theta)}(x)dx \right\} dH(\theta), \quad (29)$$

instead of (13), where the curly bracketed terms are successively the contractual indemnity, the expected audit cost, and the expected indemnity reduction.

Consider now that the burden of proof is on the insurer, and the policyholder is given the benefit of the doubt. In that case, the insurer may incur two information costs: firstly, audit cost c if signal s about the circumstances of the loss is large enough to trigger an audit; and secondly, the cost k_I (per dollar of indemnity) of providing verifiable information about x . Hence, 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 $b \in \{2, \dots, n\}$. Consequently, after an audit at stage 3, the insurer alleges misconduct $b = \hat{b}(x)$ if and only if $y_{\hat{b}(x)} \geq k_I$, and thus the equilibrium indemnity reduction is

$$z(x) = \begin{cases} 0 & \text{if } x < \tilde{x} \\ y_{\hat{b}(x)} & \text{if } x \geq \tilde{x} \end{cases} \quad (30)$$

instead of (11), where

$$\tilde{x} = \inf\{x \in (0, 1] : y_{\hat{b}(x)} \geq k_I\}. \quad (31)$$

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

$$P \geq \int_{\Theta} \pi_{b^*(\theta)} \left\{ I + \int_{0_+}^1 cE[q(s)|x]\hat{g}_{b^*(\theta)}(x)dx - I \int_{\tilde{x}}^1 (y_{\hat{b}(x)} - k_I)E[q(s)|x]\hat{g}_{b^*(\theta)}(x)dx \right\} dH(\theta) \quad (32)$$

instead of (13), where the curly bracketed terms are successively the contractual in-

demnity, the expected audit cost, and the expected indemnity reduction, net of the cost of providing verifiable information.

Comparing the two cases yields the following proposition.

Proposition 7 *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 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 $x < \tilde{x}$. Conversely, when the burden of 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 $z(x) = y_{\hat{b}(x^*)-1}$ if $x < x^*$.

This is illustrated by Figure 6 in the case $n = 4$, with $k_P = k_I$ and $x^* = \tilde{x} = x_3$. The step function in solid (black) lines corresponds to the optimal function $y_{\hat{b}(x)}$. The dotted (red) and dashed (blue) step functions correspond to function $z(\cdot)$ when the burden of proof is on the insurer and on the policyholder, respectively. Choosing $y_2 = 0$ allows us to replicate the outcome of a legal regime where the insurer has the burden of proof with one where he is given the benefit of the doubt. As the evidential cost k_P is not incurred at equilibrium when the policyholder has the burden of proof, whereas k_I is spent if $x \geq x_3$ when the insurer has to prove his allegation, we may deduce that providing the benefit of the doubt to the insurer dominates the legal regime where he has the burden of proof.

5.2 Out-of-court settlements and bad faith penalties

Most civil cases do not end up in court and one may wonder whether the possibility of out-of-court settlements affect our conclusions. Assume that the insurer has the burden of proof. Under circumstances x , he may propose to the policyholder a reduction in indemnity \hat{y} such that $0 \leq y_{\hat{b}(x)}(x) - k_I \leq \hat{y} \leq y_{\hat{b}(x)}$: in words, the insurer can credibly

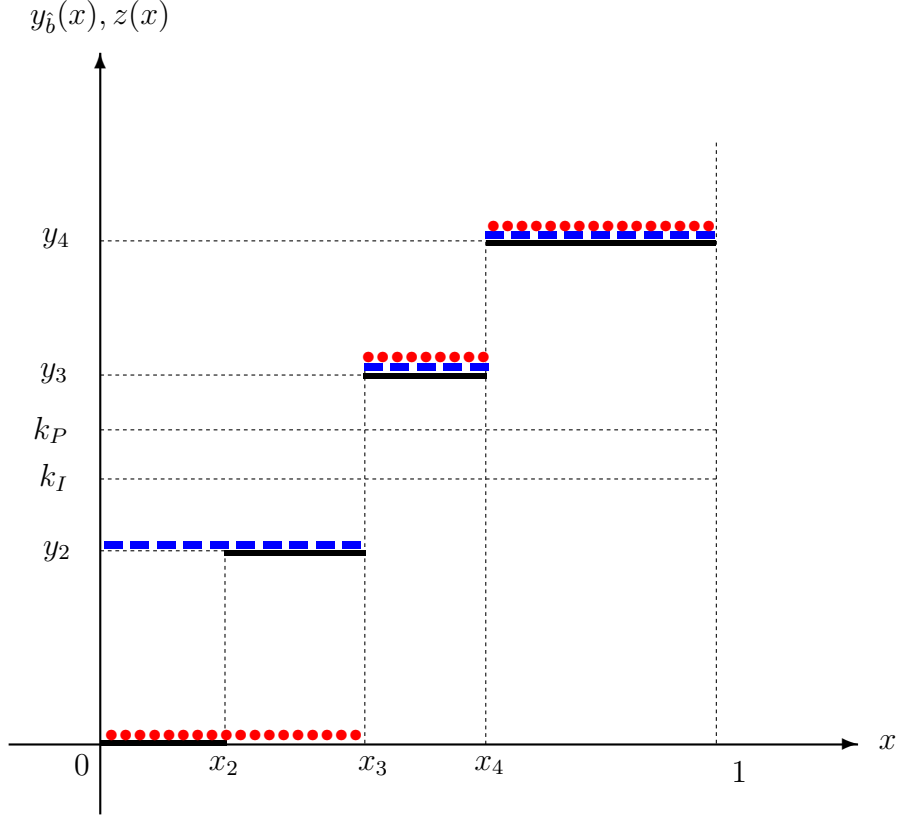


Figure 6: Burden of proof and insurance indemnity

obtain a net gain $y_{\hat{b}(x)} - k_I$ by going to court, and the out-of-court settlement would be a better deal for him and for the policyholder. Consequently, the indemnity reduction offered by the insurer and accepted by the policyholder is

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

where \tilde{x} is still defined by (31). As the insurer does not incur the cost k_I , his break-even condition is

$$P \geq \int_{\Theta} \pi_{b^*(\theta)} \left\{ I + \int_{0+}^1 cE[q(s)|x]\hat{g}_{b^*(\theta)}(x)dx - I \int_{\tilde{x}}^1 y_{\hat{b}(x)}E[q(s)|x]\hat{g}_{b^*(\theta)}(x)dx \right\} dH(\theta). \quad (33)$$

Comparing (29) and (33) with $k_P \leq k_I$, shows that giving the burden of proof to the insurer with the possibility of out-of-court settlement is equivalent to giving the burden of proof to the insured and prohibiting indemnity reductions for small values of x . It is therefore more restrictive and dominated by the solution where the burden of proof is on the insured. Consequently:

Corollary 1 *Proposition 7 remains true if the insurer can take the initiative of an out-of-court settlement when he has the burden of proof.*

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.³⁹ We could move further in that direction, and look for an optimal bad faith penalty. An hypothetical legal regime could consist in giving the burden of proof to the policyholder, with an 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 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 because her cost $k_P I$ will be reimbursed if the judge agrees with her, and consequently the insurer restricts himself to indemnity reductions 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 proof and the insurer has to pay her a sufficiently high bad faith penalty. Unfortunately, although in some cases courts may sentence 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 about 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.

6 Conclusion

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

³⁹From an empirical standpoint, insurers' bad faith penalties play a more or less important role according to the 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 breach of contract, 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.

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. 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.

Note that a difference should be made between insurance lines, depending on the degree of contract incompleteness. Automobile insurance and D&O insurance provide extreme cases in this respect, with well-established and smooth-running accident verification protocols in the first case, and a diversity of complex or opaque situations in the second. Our model justifies the fact that competition among car insurers primarily goes through the offer of rather complex contracts, including non-constant deductibles, experience rating or pay as you drive, while for D&O insurers, the question naturally arises as to whether the behavior of the insured is compatible with the coverage, which may require launching a costly audit and possibly deny coverage. These two extreme cases suggest that an inverse relationship may exist among insurance lines between contracts complexity matched with automated claim handling processes on one side, and the frequency of coverage denials based on insurance law principles on the other. Investigating how the various lines of insurance are positioned in this relationship is an empirical question that would be worth investigating.

From a more normative standpoint, our analysis also emphasizes the efficiency gains that could be derived from legal principles that would not reduce the insurer's leeway to an all-or-nothing alternative, as under the negligence rule. Adapting the diversity of penalties that may be imposed to policyholders to the diversity of possible misconducts is a way to improve the trade-off between risk-sharing and incentives under moral hazard.

These conclusions have been reached in a setting that could be easily extended in several directions. In particular, the relationship between insurer, insured and courts could also be modeled 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.⁴⁰ 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 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.⁴¹

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 involvement 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 processes coexist.

⁴⁰See Picard (2013) for a survey on the economic analysis of insurance fraud.

⁴¹As 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.
- Ambrose, J. M., Carroll, A. M., & Regan, L. (2013). The economics of liability insurance. In *Handbook of Insurance* (pp. 315–347). Springer.
- 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.
- Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1, 67–96.
- Aumann, R. J. (1987). Correlated equilibrium as expression of bayesian rationality. *Econometrica*, 55, 1–18.
- Baker, T. & Griffith, S. J. (2006). The missing monitor in corporate governance: The directors’ and officers’ liability insurer. *Georgetown Law Journal*, 95, 1795–1842.
- Bourgeon, J.-M. & Picard, P. (2014). Fraudulent claims and nitpicky insurers. *The American Economic Review*, 104(9), 2900–2917.
- Clarke, M. (1997). *The Law of Insurance Contracts*. Informa. Lloyd’s of London Press Ltd.
- 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.
- Doherty, N. A. & Schlesinger, H. (1990). Rational insurance purchasing: Consideration of contract nonperformance. *The Quarterly Journal of Economics*, 105(1), 243–253.
- 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.

- Harrison, G. W. & Ng, J. M. (2018). Welfare effects of insurance contract non-performance. *The Geneva Papers on Risk and Insurance Review*, 43(1), 39–76.
- 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.

- Spier, K. E. (2007). Chapter 4 litigation. volume 1 of *Handbook of Law and Economics* (pp. 259 – 342). Elsevier.
- 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). *The 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.
- Zou, H., Wong, S., Shum, C., Xiong, J., & Yan, J. (2008). Controlling-minority shareholder incentive conflicts and directors and officers liability insurance: Evidence from china. *Journal of Banking & Finance*, 32(12), 2636–2645.

Appendix

A Proof of Proposition 1

Let $h(\theta)$ denote the density function –or the mass of probability if any– of θ in Θ . Let $\mu_b(\theta) \geq 0$ and $\lambda \geq 0$ be Lagrange 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. $I(\theta, x)$ lead to

$$u'(W(\theta, x)) \left[1 - \sum_{b \neq b^*(\theta)} \frac{\mu_b(\theta)}{h(\theta)} \left(\frac{\pi_b}{\pi_{b^*(\theta)}} \frac{\hat{g}_b(x)}{\hat{g}_{b^*(\theta)}(x)} - 1 \right) \right] \begin{cases} \leq \lambda & \text{if } I(\theta, x) = 0, \\ = \lambda & \text{if } 0 < I(\theta, x) < L, \\ \geq \lambda & \text{if } I(\theta, x) = L, \end{cases} \quad (34)$$

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

Note that, for all $\theta \in \Theta$, the optimal solution is such that $b > b^*(\theta)$ for all $b \neq b^*(\theta)$ such that (3) is binding,⁴² and also that $\mu_b(\theta) = 0$ for all $b \neq b^*(\theta)$ such that (3) is not binding. Let

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

Hence, the l.h.s. of (34) may be written as

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

where the bracketed is decreasing w.r.t. to x if $b^*(\theta) < n$, because $\phi'_b(x) > 0$. This implies $\lambda > 0$.⁴³ Let $\theta \in \Theta$ such that $b^*(\theta) < n$. We have

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

⁴²Indeed 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.

⁴³The assumption made in footnote 19 implies $0 < I(\theta, x) < L$ in a positive-measure subset of $\Theta \times (0, 1]$, which yields $\lambda > 0$.

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

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

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

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

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

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

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

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

If $b^*(\theta) = n$, then the l.h.s. of (34) 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

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

for $b_0 \geq 2$ and $x \in (0, 1]$. When $\hat{b}(x) \geq 2$, we have

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

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

$$g_b(x') > g_b(x) \frac{g_{b_0}(x')}{g_{b_0}(x)} \quad \text{if } b > b_0,$$

$$g_b(x') < g_b(x) \frac{g_{b_0}(x')}{g_{b_0}(x)} \quad \text{if } b < b_0.$$

Hence, if $b_0 \geq 2$ and $x' > x$, we have $\Phi(b_0, x') > 0$ if $\Phi(b_0, x) > 0$. 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} .

Let $x_0, x_1 \in (0, 1], x_0 < x_1$. $\Phi(b, x)$ is continuous w.r.t. x , and thus $\Phi(\hat{b}(x_1), x_0) > 0$ when x_1 is close enough to x_0 , implying $\hat{b}(x_0) \geq \hat{b}(x_1)$. Since $\hat{b}(\cdot)$ is non-decreasing, we must have $\hat{b}(x_0) = \hat{b}(x_1)$, which implies that $\hat{b}(\cdot)$ is left-continuous.

C Proof of Lemma 2

First, we show that

Lemma 3 *When $f(\varepsilon)$ is log-concave, an increase in s shifts the probability distribution of x conditionally on s in the sense of strong FOSD.*

Proof. Let $\tilde{g}(x) = \int_{\Theta} g_{b^*(\theta)}(x) dH(\theta)$ be the density function of x for the equilibrium behavior rule $b^*(\theta)$. Using Bayes Rule and the independence of ε and x allows us to write the conditional density function of x as

$$\tilde{g}(x|s) = \frac{f(s-x)\tilde{g}(x)}{\int_0^1 f(s-u)\tilde{g}(u)du},$$

and thus the conditional c.d.f. of x is given by

$$\tilde{G}(x|s) = \frac{\int_0^x f(s-u)\tilde{g}(u)du}{\int_0^1 f(s-u)\tilde{g}(u)du}.$$

Differentiating $\tilde{G}(x|s)$ w.r.t. s shows that $\tilde{G}(x|s)$ is decreasing in s if

$$\varphi(x) \equiv \frac{\int_0^x f'(s-u)\tilde{g}(u)du}{\int_0^x f(s-u)\tilde{g}(u)du} < \frac{\int_0^1 f'(s-u)\tilde{g}(u)du}{\int_0^1 f(s-u)\tilde{g}(u)du}. \quad (35)$$

$\varphi'(x) > 0$ is a sufficient condition for (35) to hold for all $x \in [0, 1]$. This is true if

$$\frac{f'(s-x)}{f(s-x)} > \frac{\int_0^x f'(s-u)\tilde{g}(u)du}{\int_0^x f(s-u)\tilde{g}(u)du}. \quad (36)$$

Let $\eta(\varepsilon) = f'(\varepsilon)/f(\varepsilon) = d \ln f(\varepsilon)/d\varepsilon$ for all ε such that $f(\varepsilon) > 0$. (36) is rewritten as

$$\eta(s-x) > \frac{\int_0^x \eta(s-u)f(s-u)\tilde{g}(u)du}{\int_0^x f(s-u)\tilde{g}(u)du}. \quad (37)$$

Assume that $\eta'(\varepsilon) = d^2 \ln f(\varepsilon)/d\varepsilon^2 < 0$. Then we have $\eta(s-u) < \eta(s-x)$ for all $u \in [0, x)$ and (37) holds. ■

We know that, $y_b \leq y_{b+1}$ for all $b \in \{1, \dots, n-1\}$ and, furthermore, that $\hat{b}(x)$ is non-decreasing from Lemma 1. Consequently, $y_{\hat{b}(x)}$ is non-decreasing with x and Lemma 3 shows that $\mathbb{E}[y_{\hat{b}(x)}|s]$ 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 by using (9).

D Proof of Proposition 2

Assume first $c = 0$. In that case, $q(s) = 1$ for all s , i.e. $s^* = -\infty$. 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

$$(1 - \alpha) \sum_{b=2}^n \Pr(\tilde{b} = b|x_2) = \alpha \Pr(\tilde{b} = 1|x_2),$$

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

$$(1 - \alpha) \sum_{b=2}^n \frac{g_b(x_2)}{g_1(x_2)} \int_{\Theta_b^*} dH(\theta) = \alpha \int_{\Theta_1^*} dH(\theta). \quad (38)$$

From MLRP, the l.h.s. of (38) is increasing in x_2 , and it goes to 0 (resp. to $+\infty$) when x_2 goes to zero (resp. to 1) if $\int_{\Theta_1^*} dH(\theta) \neq 0, 1$. The sets $\Theta_1^*, \dots, \Theta_n^*$ depend on I, P and y , and thus (38) implicitly defines function $x_2(I, P, y) \in (0, 1)$, 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 $x \geq x_2$) in the case of an accident. The first-order optimality conditions w.r.t. I and y are written as

$$\begin{aligned}
& \int_{\Theta} \{ [u'(1)G_{b^*(\theta)}(x_2) + u'(2)(1-y)[1 - G_{b^*(\theta)}(x_2)] \\
& - \lambda [1 - y[1 - G_{b^*(\theta)}(x_2)]] \} dH(\theta) \\
& - \int_{\Theta} \{ \sum_{b \in \mathcal{B}} \mu_b(\theta) \{ u'(1)[G_b(x_2) - G_{b^*(\theta)}(x_2)] \\
& + u'(2)(1-y)[G_{b^*(\theta)}(x_2) - G_b(x_2)] \} \} dH(\theta) \\
& - \frac{\partial x_2}{\partial I} \times B \\
& \geq 0, = 0 \text{ if } I < L,
\end{aligned} \tag{39}$$

and

$$\begin{aligned}
& I[\lambda - u'(2)] \int_{\Theta} [1 - G_{b^*(\theta)}(x_2)] dH(\theta) \\
& + u'(2)I \int_{\Theta} \sum_{b \in \mathcal{B}} \mu_b(\theta) [G_{b^*(\theta)}(x_2) - G_b(x_2)] dH(\theta) \\
& - \frac{\partial x_2}{\partial y} \times B \\
& \leq 0, = 0 \text{ if } y > 0,
\end{aligned} \tag{40}$$

respectively, where

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

Suppose that $y = 0$ at an optimal solution. This implies $I < L$ and $x_2 = x_{2na}$, because $I = L$ would imply $b_{na}^*(\theta) = n$ for all 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 $B = 0$. (39) and (40) simplify to

$$\int_{\Theta} \{ (u' - \lambda) \pi_{b_{na}^*(\theta)} - u' \sum_{b \in \mathcal{B}} \mu_b(\theta) (\pi_b - \pi_{b_{na}^*(\theta)}) \} dH(\theta) = 0, \tag{41}$$

and

$$\begin{aligned}
& (\lambda - u') \int_{\Theta} [1 - G_{b_{na}^*(\theta)}(x_{2na})] dH(\theta) \\
& + u' \int_{\Theta} \sum_{b \in \mathcal{B}} \mu_b(\theta) [G_{b_{na}^*(\theta)}(x_{2na}) - G_b(x_{2na})] dH(\theta) \leq 0, \tag{42}
\end{aligned}$$

respectively. Substituting the value of $u' - \lambda$ given by (41) into (42) yields

$$\begin{aligned}
& \int_{\Theta} \sum_{b \in \mathcal{B}} \mu_b(\theta) (\pi_b - \pi_{b_{na}^*(\theta)}) dH(\theta) \times \int_{\Theta} [1 - G_{b_{na}^*(\theta)}(x_{2na})] dH(\theta) \\
& + \int_{\Theta} \pi_{b_{na}^*(\theta)} dH(\theta) \times \int_{\Theta} \sum_{b \in \mathcal{B}} \mu_b(\theta) [G_{b_{na}^*(\theta)}(x_{2na}) - G_b(x_{2na})] dH(\theta) \leq 0. \tag{43}
\end{aligned}$$

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 $b_{na}^*(\theta) = n$ almost everywhere in Θ . Furthermore, we have $b > b_{na}^*(\theta)$ and $\pi_b > \pi_{b_{na}^*(\theta)}$ if $\mu_b(\theta) > 0$ at $y = 0$ since this corresponds to the no-audit solution where the argument of footnote 42 is valid. This still holds for y positive and small enough, and implies that the first product in (43) is positive. Similarly, MLRP implies FOSD, which yields $G_b(x_{2na}) < G_{b_{na}^*(\theta)}(x_{2na})$. Since this is true for all $\theta \in \Theta$, the second product in (43) is also 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 $s^* < +\infty$ when c is not too large.

E Proof of Proposition 3

Let us consider the problem in which $\mathbb{E}u^*$ given by (12) is maximized w.r.t. $P, I, q(\cdot)$ and $z(\cdot)$, subject (14)–(16), and with $y_{\hat{b}(x)}$ replaced by $z(x) = 0$ if $x \leq \bar{x}$ and $z(x) = 1$ if $x > \bar{x}$. This corresponds to our problem under the negligence rule, with the additional condition

$$\bar{x} \in \{x_2, x_3, \dots, x_m\}, \tag{44}$$

where x_2, x_3, \dots, x_m , $m \leq n$, are the discontinuity points of $\hat{b}(x)$ in $[0, 1]$. These discontinuity points depend on the standard of proof α . Consider $\bar{x} \in [0, 1]$ as a parameter and let $\phi(\bar{x})$ be the value of $\mathbb{E}u^*$ at the relaxed problem where constraint (44) is ignored. Note that $\phi(1)$ is the expected utility $\mathbb{E}u^*$ when no audit is possible. Assume first that

$c = 0$, which gives $q(s) = 1$ for all s . In that case (10) and (13) can be rewritten as

$$u_b = (1 - \pi_b)u(W - P) + \pi_b u(W - P - L + I)\hat{G}_b(\bar{x}) + \pi_b u(W - P - L)[1 - \hat{G}_b(\bar{x})],$$

and

$$P \geq I \int_{\Theta} \pi_{b^*(\theta)} \hat{G}_{b^*(\theta)}(\bar{x}) dH(\theta),$$

respectively. Let $u_0 \equiv u(W - P - L + I)$ and $u_1 \equiv u(W - P - L)$, with $u_0 > u_1$ when (P, I) is an optimal contract for parameter \bar{x} , and let λ and $\mu_b(\theta)$ be the Lagrange multipliers corresponding to the insurer's break-even constraint and the policyholder's incentive constraint respectively when \bar{x} is close to 1. Using the envelope theorem yields

$$\begin{aligned} \phi'(\bar{x}) &= (u_0 - u_1 - \lambda I) \int_{\Theta} \pi_{b^*(\theta)} \hat{g}_{b^*(\theta)}(\bar{x}) dH(\theta) \\ &\quad + (u_0 - u_1) \int_{\Theta} \sum_{b \in \mathcal{B}} \mu_b(\theta) [\pi_{b^*(\theta)} \hat{g}_{b^*(\theta)}(\bar{x}) - \pi_b \hat{g}_b(\bar{x})] dH(\theta) \\ &\leq (u_0 - u_1) \int_{\Theta} \pi_{b^*(\theta)} g_{b^*(\theta)}(\bar{x}) \left\{ 1 + \sum_{b \in \mathcal{B}} \mu_b(\theta) [1 - \pi_b \hat{g}_b(\bar{x}) / \pi_{b^*(\theta)} \hat{g}_{b^*(\theta)}(\bar{x})] \right\} dH(\theta). \end{aligned} \quad (45)$$

The same arguments as in the proof of Proposition 2 show that, when \bar{x} is close to 1, there exists $b \in \mathcal{B}$ and a positive-measure subset of Θ in which $\mu_b(\theta) > 0$, and also that $b > b^*(\theta)$ and $\pi_b > \pi_{b^*(\theta)}$ if $\mu_b(\theta) > 0$. Using $\lim_{x \rightarrow 1} \hat{g}_{b+1}(x) / \hat{g}_b(x) \rightarrow +\infty$ when $x \rightarrow 1$ shows that the r.h.s of (45) is negative when \bar{x} is close to 1. We deduce that $\phi(\bar{x})$ is decreasing in an interval $[\bar{x}^*, 1]$ with $\bar{x}^* < 1$. As the optimal expected utility of the policyholder varies continuously with c , this remains true when c is not too large.

Let $x_2(\bar{x}, \alpha), x_3(\bar{x}, \alpha), \dots, x_{m(\bar{x}, \alpha)}(\bar{x}, \alpha)$ be the discontinuity points of function $\hat{b}(x, \alpha)$ – where we explicitly emphasize the dependence of $\hat{b}(\cdot)$ on the standard of proof α – associated to the optimal solution of the above relaxed problem with threshold \bar{x} . Here $m(\bar{x}, \alpha) \in \mathcal{B}$ denotes the largest misconduct that can be credibly alleged, i.e. $\hat{b}(x, \alpha) = m(\bar{x}, \alpha)$ for all $x \in [x_{m(\bar{x}, \alpha)}(\bar{x}, \alpha), 1]$. For all $b = 2, \dots, m(\bar{x}, \alpha)$, $x_b(\bar{x}, \alpha)$ is increasing w.r.t. α and $x_b(\bar{x}, \alpha) \rightarrow 1$ for all \bar{x} when $\alpha \rightarrow 1$. If $x_{m(\bar{x}^*, \alpha)}(\bar{x}^*, 1/2) \geq \bar{x}^*$, let $\bar{\alpha} = 1/2$. If $x_{m(\bar{x}^*, \alpha)}(\bar{x}^*, 1/2) < \bar{x}^*$, let $\bar{\alpha} > 1/2$ defined by $x_{m(\bar{x}^*, \alpha)}(\bar{x}^*, \bar{\alpha}) = \bar{x}^*$. Hence, we have $x_{m(\bar{x}^*, \alpha)}(\bar{x}^*, \alpha) \geq \bar{x}^*$ if $\alpha \geq \bar{\alpha}$. Let $\bar{b} = m(\bar{x}^*, \alpha) - 1$. Hence, under the negligence rule with standard of care \bar{b} and standard of proof $\alpha \geq \bar{\alpha}$, the indemnity is canceled if $x > x_{m(\bar{x}, \alpha)}(\bar{x}^*, \alpha)$ and it is fully paid otherwise. This is equivalent to the relaxed problem with $\bar{x} = x_{m(\bar{x}^*, \alpha)}(\bar{x}^*, \alpha) \in [\bar{x}^*, 1]$, and thus with $\phi(\bar{x}) > \phi(1)$ if $\alpha > \bar{\alpha}$.

F Proof of Proposition 4

Let us first show that $y_2 < 1$ if A is not too large. The condition $y_2 \leq 1$ is equivalently stated as $v \leq u_0 - u(W - P - L)$ which is satisfied if $v \leq u_0 - u(W - L)$, and since $v \leq A/\Delta$, if $A/\Delta \leq u_0 - u(W - L)$. As the condition $K_1(u_0, v) = \bar{W}$ defines an increasing relationship between v and u_0 , the equilibrium value of u_0 is greater than u_0^* defined by $K_1(u_0^*, 0) = \bar{W}$, i.e.

$$u^{-1}(u_0^*) \sum_{i=1,2} h_i \pi_i + u^{-1}(u_0^* + A) \left[1 - \sum_{i=1,2} h_i \pi_i \right] = \bar{W}.$$

This condition implicitly defines u_0^* as a function of A , with $u_0^{*'}(A) < 0$ and $u_0^*(0) = u(\bar{W}) > u(W - L)$. There is thus a unique $\bar{A} > 0$ satisfying $\bar{A}/\Delta = u_0^*(\bar{A}) - u(W - L)$ for given W , L and Δ , and we have $y_2 \leq 1$ at equilibrium if $A \leq \bar{A}$.

We have $v > 0$ (i.e. $y_2 > 0$) when the slope of the iso-utility lines is larger than the one of the $K_1(u_0, v) = \bar{W}$ locus at $v = 0$. Similarly, there is a corner solution with $v = A/\Delta$ (i.e. $I = L$) when the slope of the iso-utility lines is larger than the one of the $K_1(u_0, v) = \bar{W}$ locus at that point. Using

$$\frac{\partial K_1(u_0, v)}{\partial u_0} = \frac{\sum_{i=1}^2 h_i \pi_i \hat{G}_i(x_2)}{u'(u^{-1}(u_0))} + \frac{\sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2))}{u'(u^{-1}(u_0 - v))} + \frac{1 - \sum_{i=1}^2 h_i \pi_i}{u'(u^{-1}(u_0 + A - v\Delta))}$$

and

$$\frac{\partial K_1(u_0, v)}{\partial v} = -\frac{\sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2))}{u'(u^{-1}(u_0 - v))} - \Delta \frac{1 - \sum_{i=1}^2 h_i \pi_i}{u'(u^{-1}(u_0 + A - v\Delta))},$$

we have

$$\left. \frac{dv}{du_0} \right|_{\substack{K_1=\bar{W} \\ v=0}} = \frac{\sum_{i=1}^2 h_i \pi_i + (1 - \sum_{i=1}^2 h_i \pi_i) u'(u^{-1}(u_0)) / u'(u^{-1}(u_0 + A))}{\sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2)) + \Delta (1 - \sum_{i=1}^2 h_i \pi_i) u'(u^{-1}(u_0)) / u'(u^{-1}(u_0 + A))}$$

and

$$\left. \frac{dv}{du_0} \right|_{\substack{K_1=\bar{W} \\ v=A/\Delta}} = \frac{1 - \sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2)) + \sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2)) u'(u^{-1}(u_0)) / u'(u^{-1}(u_0 - v))}{\Delta (1 - \sum_{i=1}^2 h_i \pi_i) + \sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2)) u'(u^{-1}(u_0)) / u'(u^{-1}(u_0 - v))}.$$

The iso-utility lines are increasing straight lines in the (u_0, v) plane, with slope given by

$$\left. \frac{dv}{du_0} \right|_{E^{u^*=cst}} = \frac{1}{\Delta (1 - \sum_{i=1}^2 h_i \pi_i) + \sum_{i=1}^2 h_i \pi_i (1 - \hat{G}_i(x_2))}.$$

Let

$$f_0(\gamma) \equiv \frac{\sum_{i=1}^2 h_i \pi_i + \gamma(1 - \sum_{i=1}^2 h_i \pi_i)}{\gamma \Delta(1 - \sum_{i=1}^2 h_i \pi_i) + \sum_{i=1}^2 h_i \pi_i(1 - \hat{G}_i(x_2))},$$

We have $(dv/du_0)|_{K_1=\bar{W};v=0} = f_0(u'(u^{-1}(u_0))/u'(u^{-1}(u_0+A)))$ where $u'(u^{-1}(u_0))/u'(u^{-1}(u_0+A)) > 1$, while $(dv/du_0)|_{Eu^*=cst} = f_0(1)$. Hence $v > 0$ if f_0 is decreasing, i.e. if

$$\begin{aligned} 0 &> \left(1 - \sum_{i=1}^2 h_i \pi_i\right) \left[\gamma \Delta(1 - \sum_{i=1}^2 h_i \pi_i) + \sum_{i=1}^2 h_i \pi_i(1 - \hat{G}_i(x_2)) \right] \\ &- \Delta \left(1 - \sum_{i=1}^2 h_i \pi_i\right) \left[\sum_{i=1}^2 h_i \pi_i + \gamma(1 - \sum_{i=1}^2 h_i \pi_i) \right] \\ &= \left(1 - \sum_{i=1}^2 h_i \pi_i\right) \left(\sum_{i=1}^2 h_i \pi_i(1 - \hat{G}_i(x_2)) - \Delta \sum_{i=1}^2 h_i \pi_i \right) \end{aligned}$$

This is always true since we have

$$\begin{aligned} \sum_{i=1}^2 h_i \pi_i(1 - \hat{G}_i(x_2)) - \Delta \sum_{i=1}^2 h_i \pi_i &= - \sum_{i=1}^2 h_i \pi_i \left[\hat{G}_i(x_2) - \frac{\pi_2 \hat{G}_2(x_2) - \pi_1 \hat{G}_1(x_2)}{\pi_2 - \pi_1} \right] \\ &= - \frac{\pi_1 \pi_2}{\pi_2 - \pi_1} [\hat{G}_1(x_2) - \hat{G}_2(x_2)] \\ &< 0. \end{aligned}$$

Let

$$f_1(\gamma) \equiv \frac{1 + (\gamma - 1) \sum_{i=1}^2 h_i \pi_i(1 - \hat{G}_i(x_2))}{\Delta(1 - \sum_{i=1}^2 h_i \pi_i) + \gamma \sum_{i=1}^2 h_i \pi_i(1 - \hat{G}_i(x_2))},$$

we have $(dv/du_0)|_{K_1=\bar{W};v=A/\Delta} = f_1(u'(u^{-1}(u_0))/u'(u^{-1}(u_0-v)))$ where $u'(u^{-1}(u_0))/u'(u^{-1}(u_0-v)) < 1$, while $(dv/du_0)|_{Eu^*=cst} = f_1(1)$. We thus have $(dv/du_0)|_{K_1=\bar{W};v=A/\Delta} < (dv/du_0)|_{Eu^*=cst}$ if f_1 is increasing, i.e. if

$$\Delta \left(1 - \sum_{i=1,2} h_i \pi_i\right) > 1 - \sum_{i=1,2} h_i \pi_i(1 - \hat{G}_i(x_2)) = 1 - \pi_1(1 - \hat{G}_1(x_2)) - h_2(\pi_2 - \pi_1)\Delta,$$

hence if $\Delta > (1 - \pi_1(1 - \hat{G}_1(x_2)))/(1 - \pi_1)$ which may be rewritten as

$$\frac{\hat{G}_2(x_2)}{\hat{G}_1(x_2)} < \frac{\pi_1(1 - \pi_2)}{\pi_2(1 - \pi_1)} \equiv R < 1, \quad (46)$$

where $x_2 > 0$ is defined by (17). We know from MLRP, that $\hat{g}_2(x)/\hat{g}_1(x)$ is increasing for all $x \in (0, 1]$. Using $\hat{g}_2(x)/\hat{g}_1(x) \rightarrow 0$ when $x \rightarrow 0$ and $\hat{g}_2(x)/\hat{g}_1(x) \rightarrow +\infty$ when

$x \rightarrow 1$, implies $0 < x_2 < 1$. Given π_1 and π_2 , (17) defines x_2 as a decreasing function of $h_2 : x_2 = x_2(h_2)$ with $dx_2/dh_2 < 0$, $x_2(h_2) \rightarrow 1$ when $h_2 \rightarrow 0$ and $x_2(h_2) \rightarrow 0$ when $h_2 \rightarrow 1$. 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, l'Hôpital's rule gives $\lim_{x \rightarrow 0} \psi(x) = \lim_{x \rightarrow 0} \hat{g}_2(x)/\hat{g}_1(x) = 0$ and we also have $\psi(1) = 1$. Function $\psi(x)$ is thus an increasing bijection from $(0, 1]$ to itself. Condition (46) can be written as $\psi(x_2(h_2)) < R$, hence $x_2(h_2) < \psi^{-1}(R)$, i.e. $h_2 > x_2^{-1}(\psi^{-1}(R)) \equiv h^*$. We thus have $I = L$ if $h_2 \geq h^*$ and $I < L$ if $h_2 < h^*$.

G Proof of Proposition 5

Let us first show that $I < L$ at an optimal third-best solution under the negligence rule. Differentiating (22) yields

$$\frac{\partial K_2(u_0, v)}{\partial u_0} = \frac{1 - \sum_i h_i \pi_i \hat{G}_i(x_2)}{u'(u^{-1}(u_0 + A - v\Delta))} + \frac{\sum_i h_i \pi_i \hat{G}_i(x_2)}{u'(u^{-1}(u_0))}$$

and

$$\frac{\partial K_2(u_0, v)}{\partial v} = -\Delta \frac{1 - \sum_i h_i \pi_i \hat{G}_i(x_2)}{u'(u^{-1}(u_0 + A - v\Delta))}.$$

Hence, the slope of the insurer's break-even locus is given by

$$\left. \frac{dv}{du_0} \right|_{K_2=\bar{W}} = \frac{1 - [1 - u'(u^{-1}(u_0 + A - v\Delta))/u'(u^{-1}(u_0))]}{\Delta(1 - \sum_i h_i \pi_i \hat{G}_i(x_2))} \sum_i h_i \pi_i \hat{G}_i(x_2),$$

and thus

$$\left. \frac{dv}{du_0} \right|_{\substack{K_2=\bar{W} \\ v=A/\Delta}} = \frac{1}{\Delta[1 - \sum_i h_i \pi_i \hat{G}_i(x_2)]} \equiv \eta.$$

Similarly, we have

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

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

$$\Gamma(\gamma) = \frac{1 + (\gamma - 1) \sum_i h_i \pi_i [1 - \hat{G}_i(x_2)]}{\Delta [1 - \sum_i h_i \pi_i \hat{G}_i(x_2) + (\gamma - 1) \sum_i h_i \pi_i (1 - \hat{G}_i(x_2))]},$$

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

$$\Gamma'(\gamma) = -\frac{[\sum_i h_i \pi_i (1 - \hat{G}_i(x_2))] \sum_i h_i \pi_i \hat{G}_i(x_2)}{\Delta [1 - \sum_i h_i \pi_i \hat{G}_i(x_2) + (\gamma - 1) \sum_i h_i \pi_i (1 - \hat{G}_i(x_2))]^2} < 0.$$

As $u'(u^{-1}(u_0) - L) > u'(u^{-1}(u_0))$, we have $(dv/du_0)|_{Eu^*=cst, v=A/\Delta} < (dv/du_0)|_{K_2=\bar{W}, v=A/\Delta}$, which implies that $v < A/\Delta$, i.e. $I < L$ at the optimum.

Let us now show that there exists $h^{**} \in (0, 1)$ such that the optimal third-best solution with $\bar{b} = 1$ dominates the optimal no-audit solution when $h_2 < h^{**}$. Note that x_2 given by (17) is a decreasing function of h_2 , denoted $\hat{x}_2(h_2)$, with $\hat{x}_2(h_2) \rightarrow 1$ when $h_2 \rightarrow 0$ (hence, $h_1 = 1 - h_2 \rightarrow 1$). Consider x_2 and h_2 as parameters, with $\phi(x_2, h_2)$ the value function of the optimization problem where the indemnity is canceled when $x > x_2$. Note that $\phi(1, h_2)$ and $\phi(\hat{x}_2(h_2), h_2)$ correspond to the expected utility at equilibrium when there is no audit and under the negligence rule, respectively. Observe that Δ is a function of x_2 , with $\Delta'(x_2) = [\pi_1 \hat{g}_1(x_2) - \pi_2 \hat{g}_2(x_2)] / (\pi_2 - \pi_1)$. Let λ be the Lagrange multiplier associated with $K_2(u_0, v) \leq \bar{W}$. The envelop theorem yields

$$\lim_{x_2 \rightarrow 1} \frac{\partial \phi(x_2, h_2)}{\partial x_2} = \sum_i h_i \pi_i \hat{g}_i(x_2) (v - \lambda I) + v \Delta'(x_2) \left(1 - \sum_i h_i \pi_i\right) \left(\frac{\lambda}{u'(u^{-1}(\hat{u}))} - 1\right).$$

Furthermore, the first-order condition w.r.t. u_0 shows that

$$\lim_{x_2 \rightarrow 1} \lambda = \frac{1}{\frac{1 - \sum_{i=1,2} h_i \pi_i}{u'(u^{-1}(\hat{u}))} + \frac{\sum_{i=1,2} h_i \pi_i}{u'(u^{-1}(u_0))}} > u'(u^{-1}(\hat{u})).$$

since $u_0 < \hat{u}$. We deduce that

$$\frac{\partial \phi(x_2, h_2)}{\partial x_2} < v \left[\sum_i h_i \pi_i \hat{g}_i(x_2) + \Delta'(x_2) \left(1 - \sum_i h_i \pi_i\right) \left(\frac{\lambda}{u'(u^{-1}(\hat{u}))} - 1\right) \right]$$

when x_2 is close to 1. When h_2 and x_2 are close to 0 and 1 respectively, we have

$$\begin{aligned} \frac{\partial \phi(x_2, h_2)}{\partial x_2} &< v \left[\pi_1 \hat{g}_1(x_2) + \frac{\pi_1 \hat{g}_1(x_2) - \pi_2 \hat{g}_2(x_2)}{\pi_1 - \pi_2} (1 - \pi_1) \left(\frac{\lambda}{u'(u^{-1}(\hat{u}))} - 1 \right) \right] \\ &= \frac{v \pi_1 \hat{g}_1(x_2)}{\pi_1 - \pi_2} \left[\pi_1 - \pi_2 + \left(1 - \frac{\pi_2 \hat{g}_2(x_2)}{\pi_1 \hat{g}_1(x_2)} \right) (1 - \pi_1) \left(\frac{\lambda}{u'(u^{-1}(\hat{u}))} - 1 \right) \right] \\ &< 0, \end{aligned}$$

since $\hat{g}_2(x_2)/\hat{g}_1(x_2) \rightarrow +\infty$ when $x_2 \rightarrow 1$. Since $\hat{x}_2(h_2) < 1$ and $\hat{x}_2(h_2) \rightarrow 1$ when $h_2 \rightarrow 0$, we deduce $\phi(\hat{x}_2(h_2), h_2) > \phi(1, h_2)$ when h_2 is close to 0, i.e. $h_2 < h^{**} \in (0, 1)$.

H Monotonicity of $\Delta_2(x)/\Delta_1(x)$

Considering (23) as a system of equations with unknowns \underline{v} and \bar{v} yields

$$\underline{v} = \frac{(A_1 - z)\Delta_2(x_3) - (A_2 - z)\Delta_1(x_3)}{\Delta_1(x_2)\Delta_2(x_3) - \Delta_1(x_3)\Delta_2(x_2)}, \bar{v} = \frac{(A_2 - z)\Delta_1(x_2) - (A_1 - z)\Delta_2(x_2)}{\Delta_1(x_2)\Delta_2(x_3) - \Delta_1(x_3)\Delta_2(x_2)} \quad (47)$$

with $z = 0$ when $I = L$. We have to verify that the expressions given by (47) are positive, i.e. that we have

$$\Delta_2(x_3)/\Delta_1(x_3) > (A_2 - z)/(A_1 - z) > \Delta_2(x_2)/\Delta_1(x_2). \quad (48)$$

The condition $\Delta_2(x_3)/\Delta_1(x_3) > \Delta_2(x_2)/\Delta_1(x_2)$ can be rewritten as

$$\int_{x_2}^{x_3} \left(\frac{d \ln \Delta_2(x)}{dx} - \frac{d \ln \Delta_1(x)}{dx} \right) dx > 0,$$

which is satisfied if

$$\frac{d \ln \Delta_2(x)}{dx} > \frac{d \ln \Delta_1(x)}{dx} \quad (49)$$

for all $x \in [0, 1]$. Denote $\beta_i(x) \equiv \pi_{i+1}[1 - \hat{G}_{i+1}(x)] - \pi_i[1 - \hat{G}_i(x)]$. We have $\Delta_i(x) = \beta_i(x)/\beta_i(0)$ and

$$\begin{aligned} \frac{d \ln \Delta_i(x)}{dx} &= \frac{\Delta'_i(x)}{\Delta_i(x)} = \frac{\pi_i \hat{g}_i(x) - \pi_{i+1} \hat{g}_{i+1}(x)}{\pi_{i+1}[1 - \hat{G}_{i+1}(x)] - \pi_i[1 - \hat{G}_i(x)]} = \frac{\beta'_i(x)}{\beta_i(x)} \\ &\equiv -\gamma_i(x) \end{aligned} \quad (50)$$

As $\beta_i(0) = \pi_{i+1} - \pi_i$, we get

$$\beta_i(x) = (\pi_{i+1} - \pi_i) \exp \left\{ - \int_0^x \gamma_i(t) dt \right\}.$$

Moreover, as $\beta_i(1) = 0$, we must have

$$\lim_{x \rightarrow 1} \int_0^x \gamma_i(t) dt = +\infty. \quad (51)$$

Condition (49) is thus given by

$$\gamma_2(x) < \gamma_1(x), \quad (52)$$

for all $x \in [0, 1]$. Denote

$$\xi_i(x) = \pi_i [1 - \hat{G}_i(x)],$$

for $i = 1, 2, 3$. Given π_1 and \hat{G}_1 , as $\beta_i(x) = \xi_{i+1}(x) - \xi_i(x)$, we have $\xi_2(x) = \xi_1(x) + \beta_1(x)$ and $\xi_3(x) = \xi_2(x) + \beta_2(x)$. Given π_2 and π_3 , we can deduce \hat{G}_2 and \hat{G}_3 . Also, using (50) and $\lim_{x \rightarrow 0} \Delta_i(x) = 2$, we have

$$\Delta_i(x) = 2 \exp \left\{ - \int_0^x \gamma_i(t) dt \right\},$$

and thus $\Delta_2(x) > \Delta_1(x)$ for all $x \in (0, 1)$ under (52) and $\Delta_i'(x) = -\gamma_i(x)\Delta_i(x) < 0$ if $\gamma_i(x) > 0$ for all x . These probability distribution must satisfy MLRP, which can be checked on specific examples.

Example: Consider $\gamma_i(x) = a_i/(1-x)$, $a_i > 0$ to satisfy condition (51), and suppose \hat{G}_1 is uniform, i.e. $\hat{g}_1(x) = 1$ for all x . We have $\xi_{i+1}'(x)/\xi_i'(x) = \pi_{i+1}\hat{g}_{i+1}(x)/\pi_i\hat{g}_i(x)$ and thus MLRP is satisfied if $\xi_{i+1}'(x)/\xi_i'(x)$ is increasing. We have

$$\frac{\xi_2'(x)}{\xi_1'(x)} = \frac{\xi_1'(x) + \beta_1'(x)}{\xi_1'(x)} = 1 - \frac{\beta_1'(x)}{\pi_1},$$

since $\xi_1(x) = \pi_1(1-x)$. Using $\beta_1'(x) = -\gamma_1(x)\beta_1(x)$, we get

$$\frac{\xi_2'(x)}{\xi_1'(x)} = 1 + \gamma_1(x)\beta_1(x)/\pi_1.$$

MLRP implies that $\gamma_1(x)\beta_1(x)$ is increasing, hence $\gamma_1'(x) - \gamma_1^2(x) = a_1(1-a_1)/(1-$

$x)^2 > 0$ for all x , i.e. $a_1 < 1$. We also have

$$\begin{aligned}\frac{\xi_3'(x)}{\xi_2'(x)} &= \frac{\xi_2'(x) + \beta_2'(x)}{\xi_2'(x)} = 1 + \frac{\beta_2'(x)}{\xi_1'(x) + \beta_1'(x)} \\ &= 1 + \frac{\gamma_2(x)\beta_2(x)}{\pi_1 + \gamma_1(x)\beta_1(x)},\end{aligned}$$

which is increasing if

$$\begin{aligned}0 &< [\gamma_2'(x)\beta_2(x) + \gamma_2(x)\beta_2'(x)][\pi_1 + \gamma_1(x)\beta_1(x)] - [\gamma_1'(x)\beta_1(x) + \gamma_1(x)\beta_1'(x)]\gamma_2(x)\beta_2(x) \\ &= a_2(1 - a_2)\pi_1\beta_2(x)/(1 - x)^2 + (a_1 - a_2)a_1a_2\beta_1(x)\beta_2(x)/(1 - x)^3,\end{aligned}$$

where the first term is positive if $a_2 < 1$, and the second if $a_1 > a_2$. MLRP is thus satisfied if $1 > a_1 > a_2 > 0$. As a consequence, we have $\Delta_2(x_3)/\Delta_2(x_2) > \Delta_1(x_3)/\Delta_1(x_2)$ and $\Delta_2(x) > \Delta_1(x)$ for all $x \in (0, 1)$.

I Proof of Proposition 6

We have

$$z = u(w - P) - u(w - P - L + I) \leq (L - I)u'(w - P - L + I),$$

since $u'' < 0$ and $I \leq L$. Moreover, with $\underline{v} \geq 0$ and $\bar{v} \geq 0$ (and $c = 0$) we have $P \leq \bar{\pi}I$ where $\bar{\pi} = \sum_{i=1}^3 h_i\pi_i$, implying $u'(w - P - L + I) < u'(w - L + (1 - \bar{\pi})I)$. Consequently,

$$0 \leq z \leq (L - I)u'(w - L + (1 - \bar{\pi})I),$$

where the right hand term decreases with I . We thus have

$$0 \leq z \leq Lu'(w - L) \equiv \hat{z}.$$

Assume that $\hat{z} < \inf\{A_1, A_2\}$. If $A_2 > A_1$, the ratio $(A_2 - z)/(A_1 - z)$ increases with $z \in [0, \hat{z}]$ and (48) is satisfied whatever the optimal value of z if

$$\frac{\Delta_2(x_3)}{\Delta_1(x_3)} > \frac{A_2 - \hat{z}}{A_1 - \hat{z}} \quad \text{and} \quad \frac{\Delta_2(x_2)}{\Delta_1(x_2)} < \frac{A_2}{A_1}.$$

These conditions hold if A_1 and A_2 satisfy

$$\frac{\Delta_2(x_3)}{\Delta_1(x_3)} > \frac{A_2}{A_1} > \frac{\Delta_2(x_2)}{\Delta_1(x_2)}, \quad (53)$$

and \hat{z} is small enough. If $A_2 < A_1$, the ratio $(A_2 - z)/(A_1 - z)$ decreases with $z \in [0, \hat{z}]$ and (48) is satisfied whatever the optimal value of z if

$$\frac{\Delta_2(x_3)}{\Delta_1(x_3)} > \frac{A_2}{A_1} \quad \text{and} \quad \frac{\Delta_2(x_2)}{\Delta_1(x_2)} < \frac{A_2 - \hat{z}}{A_1 - \hat{z}},$$

hence if A_1 and A_2 satisfy (53) and \hat{z} is small enough. In the example, we have $\Delta_2(x)/\Delta_1(x) = 1/(1-x)^{a_1-a_2}$ with $a_1 > a_2$ and thus $\lim_{x \rightarrow 1} \Delta_2(x)/\Delta_1(x) = +\infty$ while we have $\lim_{x \rightarrow 0} \Delta_2(x)/\Delta_1(x) = 1$.

J Proof of proposition 7

In the fourth-best case, an allocation a is defined by a contract (P, I) , an audit strategy $q(\cdot)$ with threshold s^* , insurance law rules $\{y_b, b \in \mathcal{B}\}$, a behavioral rule $b^*(\cdot)$ and a (potential) allegation function $\hat{b}(\cdot)$. The indemnity reduction function $z(x)$ is defined by (28) or (30) instead of (11). 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), (12), (14)-(16), and (30)-(32). 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 $x < \tilde{x}$ w.l.g.

Let \mathcal{A}_P be 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), (12), (14)-(16), and (27)-(29). Consider first the case where $k_P = k_I$. We have $x^* = \inf\{x \in (0, 1] : y_{\hat{b}(x)} \geq k_P\} = \tilde{x}$, the indemnity reduction functions (28) and (30) are the same, and since a_I satisfies (32), it also satisfies (29) strictly. Hence $a_I \in \mathcal{A}_P$. However, since (29) is satisfied but not binding, a_I is not the optimal allocation when the burden of 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 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.