CESifo WORKING PAPERS

6669
2017
September 2017

Nitpicky Insurers and the Law of Contracts

Jean-Marc Bourgeon, Pierre Picard

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6669*
Category 2: Public Choice

Nitpicky Insurers and the Law of Contracts

Abstract

The standard economic analysis of the insured-insurer relationship under moral hazard postulates a simplistic setup that hardly explains the many features of an insurance contract. We extend this setup to include the situation that the insured was facing at the time of the accident and the circumstances of the loss. We show that if this information is costlessly observable, then it should be included in the contract to improve the risk sharing-incentive trade-off under moral hazard. However, in practice the insurer observes the circumstances of the loss only in particular cases - most of the time by performing a costly audit - and almost never the situation the insured was facing at the time of the accident. The resulting incompleteness of the contract opens the door to controversies and disputes that may lead to judicial procedures. We show how the law of insurance contracts should allow insurers to incentivize policyholders to exert an adequate level of effort, and, at the same time, to limit their propencity to nitpick.

JEL-Codes: D820, D860, G220.

Keywords: insurance, moral hazard, incomplete contracts.

Jean-Marc Bourgeon
INRA and CREST
Ecole Polytechnique
UMR Économie Publique
16 rue Claude Bernard
France – 75231 Paris Cedex 05
bourgeon@agroparistech.fr

Pierre Picard
CREST – Ecole Polytechnique
Department of Economics
France - 91128 Palaiseau Cedex
pierre.picard@polytechnique.edu

September 12, 2017
Pierre Picard acknowledges financial support from Investissements d’Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

1 Introduction

It is striking to observe how an average citizen's perception of what an insurance contract is often differs from its usual definition by economists. The counterpart of the insurance premium as perceived by policyholders is generally nothing more than an imprecise expectation of what future indemnities will be, should a loss occur. This is particularly true for lines of risk where, rightly or wrongly, policyholders think that insurers have leeway in settling claims, and may nitpick if they believe it possible, on the indemnity payment. On the other hand, economic analysis commonly relies on a crude description of what an insurance policy is, without room for ambiguity. An insurance contract is simply characterized by the indemnity schedule that defines the insurance payment depending on the policyholder's loss, by the premium charged by the insurer, and sometimes by a policy dividend rule.[1] Defining an insurance policy this way is suitable for analyzing risk-sharing and competitive interactions in a large variety of contexts, whether they be characterized by transaction costs, information asymmetry, parameter uncertainty, limits to risk pooling or other features.

Nevertheless, this approach imperfectly reflects actual insurance mechanisms. In particular, the indemnity payment may not only depend on the financial assessment of the policyholder's loss, but also on the circumstances under which this loss occurred. This includes the causal mechanism that links the operative event with the damage, and the qualitative description of the loss itself, in particular through exclusions stated in the small print of the contract,[2] and the soft-law guidelines provided by market regulators, but also, more indirectly, when the insurer has some leeway to challenge the legitimacy of the claim by invoking provisions or practices of insurance law.

Conditioning the insurance payment on the circumstances of the loss leads to a lower predictability of the coverage in the case of an accident. Yet, from a risk-sharing standpoint, there is no reason why this should be so. However, restricting coverage according to the circumstances of the loss may be worthwhile under moral hazard if circumstances are informative concerning the policyholder's effort, whether it be associated with risk prevention or loss reduction. This is

[1] Of course, in practice, these ingredients of insurance contracts take many different forms, including, for instance, experience rating in automobile insurance, or fee-for-service payment in health insurance.

[2] For instance, a corporate property policy may exclude the damage resulting from fire caused by an explosion, or from the transportation of hazardous materials. Similarly, a homeowner policy may exclude theft that may result from a lack of minimum precaution, such as locking the door when leaving home.

a well-known result in incentive theory (in particular Holmström, 1979), but its implications for the design of insurance contracts may not have been sufficiently scrutinized. The objective of this paper is to explore these issues.

To do so, we will first consider the benchmark case where the behavior of the policyholder is not observed by the insurer, but both the situation in which the policyholder found herself while deciding on her behavior and the circumstances of the losses that may arise are verifiable facts upon which the contractual insurance coverage can be conditioned.[3] We will show that, in such a setting, the insurer should pay full compensation or entirely deny the claim when the circumstances are very favorable or very unfavorable (i.e. when they most likely correspond, respectively, to a high or low effort level), and he should provide partial coverage in more ambiguous cases. Furthermore, the optimal coverage schedule also depends on the situation in which the policyholder found herself when she chose her more or less risky behavior. In other words, unfavorable accident circumstances may induce a more or less severe indemnity cut according to the situation that was experienced by the policyholder.

Although the small print of insurance policies and soft-law guidelines may produce links between the circumstances of an accident, the situation the policyholder found herself in and the actual coverage, it remains no less true that, quite often, the practice of insurance contracting goes beyond clauses of contracts and regulators' recommendations. Insurers have leeway in determining the amount of compensation policyholders receive and do more than simply enforce exclusion clauses, soft-law instruments, or taking account of force majeure.

This refers to the conflicted dimension of the insurer-insured interaction which will be the core of our analysis. This confrontational situation is fundamentally related to the incompleteness of insurance contracts. Exclusion clauses, force majeure stipulations and regulation guidelines are indeed just the tip of the iceberg when it comes to describing how the insurance coverage should

[3]Beyond what is written in the small print of an insurance policy, the coverage may be conditioned upon the circumstances of the accident when the treatment of information is regulated by well-formatted processes (for instance, police and expert reports for car accidents) and all the possible circumstances are classified in a comprehensive way by insurance regulators, so that the interpretation of contracts precludes any ambiguity. This is the case in risk line such as automobile insurance or homeowner insurance, where many similar accidents occur each year, and market regulation has established a well-defined interpretation of circumstances with little room for ambiguity. A force majeure clause is an example of a feature that takes into account the situation the policyholder may face when she chooses her action.

be ideally adapted to the diversity of the circumstances that characterize insurable losses and the situations that lead policyholders to behave in a particular way. More often than not, this diversity cannot be fully taken into account in insurance contracts and in soft law instruments because it is so diverse that all possible cases cannot be exhaustively listed, and because what really occurred is, at least partially, private information of the claimant.

Observing the detailed circumstances of a loss requires a costly state verification process, but such an audit is justified and beneficial only when the insurer has reason to believe that the policyholder misbehaved in some way. Furthermore, apart from particular cases where force majeure is related to a publicly observable event (such as a well-documented natural disaster), the situation that led the policyholder to behave in a certain way cannot be verified by the insurer, which makes it irrelevant to the insurance coverage. In short, in a context of moral hazard, insurance contracts are frequently incomplete, in the sense that they do not condition the coverage on all relevant information.

This does not necessarily mean that the link between the indemnity paid to the claimant and the circumstances of the loss no longer exists. The argument developed in this paper is that this link may go through legal disputes, in which the insurer invokes legal means to fully or partially deny the claim. His denial of coverage may be dismissed in court if he reports information on circumstances that do not provide enough support for his allegation.[4] Our approach will consist in assuming that the insurer may either validate claims on the basis of the (freely available) soft information, or search for verifiable information through a costly audit in order to sustain a nitpicking strategy that would not be dismissed in court. We will consider a simple setting where the standard of proof used by courts is the balance of probabilities, i.e. courts weigh up the evidence based on the circumstances of the loss and decide which was the most probable behavior of the policyholder, and ultimately whether the indemnity should be paid fully or not.[5] In such a setting, allowing insurers to allege possible misconduct of the policyholder is an indirect way to condition the insurance

[4] If well-established case law does not provide an immediate answer to a dispute, the ultimate decisions of judicial authorities are based on fundamental (not case-specific) principles of the law of insurance contracts: duty of utmost good faith, prohibition of preexisting conditions of non-disclosure, cancellation of contract for negligence, invalidity of a claim because of misrepresentation, absence of evidence about the operative event, onus of the proof regarding the loss, just to name a few.

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

coverage on the circumstances of the loss, because these circumstances affect the final decision that will result from the application of the law of contracts.

The relationship between contract law and economic efficiency when contracts are incomplete has been extensively discussed in the literature on incentives and contracts. The starting point is in the theory of the firm and of corporate finance and, more specifically, in the under or over-investment problems that arise when the state-contingent sharing of surplus cannot be exhaustively described in contractual arrangements between stakeholders. A major trend in the literature has put the emphasis on how the allocation of property rights and incentive mechanisms could restore economic efficiency when complete contracts are unfeasible.[6] Another approach, more in line with the perspective of the law and economics, puts the emphasis on how legal rules restrict the set of feasible contracts and constrain the process of adversarial litigation in contract enforcement. Instead of explicitly distinguishing what is verifiable by courts and what is not, it consists in analyzing incomplete contracts as agreements that do not specify what should be done by the parties in some contingencies and that include references to broad standards.[7,8] Such a form of incompleteness is usually rationalized by arguing that shifting some decisions to the back end (i.e., to the adversarial stage) reduces front-end transaction costs. The compliance or non-compliance with the standards is appreciated by courts, with a relative rather than absolute evaluation of evidence, like the balance of probabilities. Our analysis is linked to this second approach by considering a setting where insurers may refer to behavioral standards to deny claims, and by focusing attention on how insurance law

[6] A large trend of this litterature focuses on the case where some information is missing at the ex ante stage and revealed ex post - see Nöldeke & Schmidt (1995) and Edlin & Hermalin (2000) in the case of option contracts. Of particular interest is how contracting parties try to regulate the renegotiation process when information is revealed, in order to predetermine the sharing of the surplus - see Grossman & Hart (1986); Hart & Moore (1990); Aghion et al. (1994); Edlin (1996); Edlin & Reichelstein (1996).

[7] Scott & Triantis (2005) 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.

[8] See for instance Shavell (1980) on damage measures for breach of contract. He shows that a moderate damage measure is desirable because it induces performances if and only if the cost of performance is relatively low, and for that reason it acts as a substitute for complete contracts. Another example is provided by Ayres & Gertner (1989) on default rules, i.e., rules that govern economic relationships unless the parties contract around them. They analyze how lawmakers may reduce strategic rent-seeking by establishing penalty defaults that encourage the better informed parties to reveal their information. Note that insurance contracts are full of references to broad standards. For instance, reasonable care and good faith are commonplace in insurance contracts, although the fact that a policyholder exercized reasonable care or was in good faith is a matter of appreciation within the expertise of courts.

and adversarial litigation restrict their discretionary power.

In a previous paper related to the insurance fraud issue (Bourgeon & Picard, 2014), we showed that allowing insurers to cut the indemnity according to the circumstances of the loss could be welfare improving, when such a nitpicking strategy acts as an incentive device to perform costly audits. We assumed that the law of insurance contracts restricts the contractually feasible indemnity cuts, but the determinants of these legal stipulations were not analyzed. The present paper goes a step further by analyzing how the law of contracts may be an indirect way to condition insurance coverage on the circumstances of the loss, thereby improving the incentive-risk sharing trade-off that is inherent in insurance under moral hazard.

The rest of the paper is as follows. Section 2 introduces the model. Section 3 considers the benchmark moral hazard problem, in which the contractual insurance indemnity is specified for all possible contingencies that may be at the origin of a loss, and all conceivable circumstances of the loss. Section 4 raises the issue of incomplete insurance contracts when gathering information on circumstances requires going through a costly audit, and the contingencies at the origin of a loss cannot be contracted for. It shows how the law of contracts allows society to arbitrate between the insured's lack of desire to exert an effort and the insurer's propensity to nitpick. Section 5 concludes. Proofs are in an appendix.

2 The model

Consider an insurance company providing coverage to a risk-averse individual (household or firm) against accidents that may result in a loss L. The occurrence of the loss depends on the policyholder's behaviors which is indexed by $b \in \mathcal{B} = \{1, 2, \ldots, n\}$ and ranks the probability of accident π_b increasingly, i.e. $\pi_1 < \pi_2 < ... < \pi_n$. Hence, $b = 1$ corresponds to a cautious behavior with the lowest probability of accident, and the other behaviors $b \in \{2, ..., n\}$ correspond to various types of misconducts, increasingly risky, but also decreasingly demanding in terms of effort.[9] The disutility of 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. We assume $\theta \in \Theta$, where Θ is the (mul-

[9] For instance, a car driver may exert a low level of effort because she does not adequately maintain her vehicle, or because her speed is not appropriate, or because she drives after drinking, or because of a mixture of these behaviors.

tidimensional) set of possible states, and the disutility of behavior b in state θ is denoted $d_b(\theta)$, with $d_b(\theta) > d_{b+1}(\theta)$, for all $b \in \mathcal{B}$ and all $\theta \in \Theta$. Hence, in all states, less risky behaviors entail a larger disutility because they require more effort on the part of the policyholder. We assume that θ is distributed in Θ according to a continuous c.d.f. $H(\theta)$ with density function $h(\theta)$. If the policyholder chooses behavior b in state θ, then her utility is $u(W_f) - d_b(\theta)$ where W_f is her final wealth and u is a (twice continuously differentiable) von Neumann-Morgenstern utility function such that $u' > 0, u'' < 0$.

We denote by $b^\star(\theta)$ the behavior chosen by the policyholder in state θ, a choice that is governed by the specification of the insurance contract. In particular, an insurance policy inducing a behavior $b^\star(\theta) = 1$ for all θ is usually suboptimal because $d_1(\theta)$ may be very large in some states θ.[10]

In addition to the situations (represented by state θ) that have led the policyholder to engage in some behavior, any accident that may occur is also characterized by its circumstances, i.e., by all the specificities of the operative event at the origin of the loss. These circumstances are denoted $\omega \in \Omega$, with $(\Omega, \mathcal{F}, \mathbb{P}_b)$ a probability space where the probability measure $\mathbb{P}_b$ depends on the policyholder's behavior.

We know from Milgrom (1982) that we can associate a real variable x to circumstances ω through a function $x = \varphi(\omega)$, with $\varphi : \Omega \longrightarrow [0, 1]$, such that x is a sufficient statistic for b and satisfies the strict Monotone Likelihood Ratio Property (MLRP). More explicitly, the likelihood ratios $\phi_b(x) \equiv g_{b+1}(x)/g_b(x)$ are increasing over $[0, 1]$ for all $b \in \{1, \ldots, n-1\}$.[11] Function φ is not defined in a unique way, and in what follows, this function is taken as given and representing the insurer's information system, including all relevant information about accidents that the insurer is able to process, either costlessly in section 3 or after auditing the claim in section 4. For the sake of simplicity and by an abuse of language, in what follows we may refer to x as the circumstances of the accident, although this is only the information about these circumstances that are available to the insurer.

[10] To take an extreme example, think of a man who breaks the speed limit when driving his wife to the hospital maternity, or, less dramatically, think of a driver who is worrying about arriving late at an important business meeting.

[11] In particular, for any nondegenerate prior on b, an increase in x induces a FOSD shifts in the posterior probability distribution of b. In this sense, a larger x can be interpreted as a "bad news", i.e., as suggesting that the behavioral parameter b was high.

3 Information transparency

Information is said to be transparent when the insurer costlessly observes both the ex ante events θ that have conditioned the policyholder's behavior, and the ex post accident circumstances x. The insurer can condition coverage upon θ and x, but not upon the unobservable policyholder's actual behavior b. Hence, an insurance contract specifies a premium P paid at the outset, and an indemnity $I(\theta, x)$ for all $(\theta, x) \in \Theta \times [0,1]$ in the case of a loss.[12]

With such a contract, the policyholder's final wealth is $W_f = W - P$ if there is no accident, and $W_f = W - P - L + I(\theta, x)$ in the case of an accident. Thus, conditionally on state $\theta \in \Theta$, the expected utility of a policyholder with behavior $b \in \mathcal{B}$ is written as $u_b(\theta) - d_b(\theta)$, where

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

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

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

and

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

respectively. We neglect any transaction costs, and assume that insurers are risk neutral. Hence, the insurance premium must at least cover the expected indemnity payments, i.e.

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

Finally, we 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. Hence, taking into account the non-negativity constraint on coverage, the indemnity schedule is such that

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

[12] Circumstances x correspond to objective facts that characterize the operative event at the origin of the loss, while θ refers to everything that conditions the policyholder's decision making about her behaviour b. Assuming that θ is observed by the insurer and that it can condition the insurance compensation is thus a very strong assumption that serves as a benchmark for the more realistic setup investigated in section 4.

The optimal insurance contract maximizes $Eu^\star$ given by (2) with respect to P, $I(\cdot)$ and $b^\star(\cdot)$ subject to (3),(4) and (5). $I^*(\theta,x)$ denotes the optimal indemnity schedule, and not surprisingly, $I^*(\theta,x)$ actually depends on θ and x. Indeed, for a given behavioral rule $b^*(\theta)$, the uncertainty about θ should be taken into account in the insurance coverage.[13] Furthermore, x is informative about the policyholder's effort in the sense of Holmström (1979) and, in our moral hazard context, it should condition the transfer from insurer to policyholder. The remainder of this section shows how this conditioning should be implemented.

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

$$\begin{array}{ll} I^\star(\theta,x) = L & \textit{if } 0 \le x < \underline{x}(\theta) \ \textit{if } \underline{x}(\theta) > 0, \\ \begin{array}{c} 0 < I^\star(\theta,x) < L \\ dI^\star(\theta,x)/dx < 0 \end{array} & \textit{if } \ \underline{x}(\theta) < x < \overline{x}(\theta), \\ I^\star(\theta,x) = 0 & \textit{if } \overline{x}(\theta) < x \le 1 \ \textit{if } \overline{x}(\theta) < 1. \end{array}$$

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

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

Figure 1

As a comparison, we may consider the two polar cases where circumstances x are either totally uninformative or perfectly informative about behavior b. Circumstances do not convey information

[13] By way of follow-up to footnote 8, it is conceivable that, when a road traffic offense has been committed, cutting the indemnity is less recommended in the case of the man who drives his wife to the maternity in emergency, than when the same person is worried about arriving late at a business meeting. The behavioral disutility associated with obeying speed limits is probably larger in the first case than in the second.

on the policyholder's conduct when $g_b(x) = 1$ for all $x \in [0,1]$ and all $b \in \mathcal{B}$. In this case, the optimal insurance indemnity does not depend on x, and partial coverage is optimal when the policyholder should be incentivized to exert effort. In other words, we should have $I^*(\theta, x) = \overline{I}(\theta) \in (0, L]$, with $\overline{I}(\theta) < L$ if $b^*(\theta) < n$.[14] This corresponds to the most common setting of insurance contracts under moral hazard. Conversely, signal x is perfectly informative when the policyholder's behavior can be deduced from the circumstances of the loss. This is the case if there is a partition of interval $[0,1]$, each subinterval being associated with a particular behavior b. More explicitly, there exists a sequence $\{\widetilde{x}_b, b \in \mathcal{B}\}$, with $\widetilde{x}_1 = 0, \widetilde{x}_n = 1$ and $\widetilde{x}_b < \widetilde{x}_{b+1}$, such that $g_b(x) > 0$ if $x \in [\widetilde{x}_b, \widetilde{x}_{b+1})$ and $g_b(x) = 0$ otherwise. In that case, in state θ, the policyholder is incentivized to choose behavior $b^*(\theta)$ with $I^*(\theta, x) = L$ if $x \in [0, \widetilde{x}_{b^*(\theta)+1})$ and $I^*(\theta, x) = 0$ if $x \in [\widetilde{x}_{b^*(\theta)+1}, 1]$. With such a stepwise decreasing indemnity schedule, the policyholder chooses $b = b^*(\theta)$ in state θ with full coverage of her losses. If the circumstances of the loss are partially informative about the policyholder's conduct, then the optimal indemnity schedule reaches a compromise between these two polar cases: it is continuously decreasing, possibly with full coverage and exclusions under, respectively, the most favorable and unfavorable circumstances.

In practice, conditioning insurance coverage on the circumstances of the loss frequently goes through exclusions. The insurer commits to paying an indemnity $\overline{I}(\theta)$ in the case of an accident, except under well-defined circumstances of the loss. This amounts to restricting the indemnity schedule to

$$I(\theta, x) = \begin{cases} \overline{I}(\theta) \text{ if } x \notin X(\theta) \\ 0 \text{ if } x \in X(\theta), \end{cases}$$

where $X(\theta) \subset [0,1]$ is the set of signals for which an exclusion applies.

Proposition 2 *The optimal exclusion-based insurance contract is such that* $X(\theta) = [\hat{x}(\theta), 1]$ *with* $\hat{x}(\theta) < 1$ *if* $b^*(\theta) < n$.

Proposition 2 characterizes the optimal solution when the insurance contract just includes a fixed indemnity $\overline{I}(\theta)$ and exclusions. Unsurprisingly, an exclusion applies when x is larger than a threshold $\hat{x}(\theta)$, that is to say, under the most unfavorable circumstances, and this exclusion clause is effective for all states θ such that $b^*(\theta) < n$, meaning that some effort is required.

[14] In cases of force majeure, not exerting effort may be optimal, but these are truly exceptional cases.

4 Costly state verification and incomplete contracts

The previous setup does not realistically reflect the business practice of insurers when it comes to paying benefits to policyholders. Indeed, except in exceptional cases the state in which policyholders find themselves is not observed by insurers and, furthermore, obtaining information about the circumstances of the loss 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 of the loss. This information is obtained only for the subset of claims that are subject to a costly audit, and, to guarantee equal treatment for policyholders, it will not be used to enforce contractual clauses of the policies. Hence, in most cases encountered in practice, 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.

In fact, when the circumstances of the loss are informative about the policyholder's behavior, auditing the claim may be the starting point of disputes between insurer and insured, that may end with an amicable settlement, by resorting to an arbitrator or by going to court. In some way or another, these disputes are resolved in compliance with the law of contracts. This is the interaction between the verification of the circumstances of the loss and the stipulations of insurance law that we will contemplate in what follows. In essence, after observing a signal correlated with the circumstances of a loss, the insurer may decide to gather verifiable evidence about the behavior of the insured. This information may allow him to invoke the law of insurance contracts in order to justify a cut in the indemnity.

The insurer is willing to perform an expensive audit process (checks and, possibly, testimonials, invoices, an inventory, etc.) if it reduces the expected indemnity payment by an amount larger than the audit cost. In view of the evidence, the insurer may be allowed by law to wholly or partially cancel the contractual compensation. In concrete terms, the insurer may allege that the policyholder misbehaved in some way and may invoke a legal means under the form of a broad standard that justifies the cut in indemnity. For instance, in many property insurance settings the law of contracts specifies that the policyholder has a duty of care, and the insurer is allowed by law to totally or partially reject the claim if he considers that this principle was disregarded, and if courts do not invalidate his decision. For the sake of illustration, Table 1 provides examples of

insurance lines	misconducts	legal means
health insurance	hiding smoker status	non-disclosure
life insurance	misrepresenting pre-existing heath conditions	misrepresentation
homeowner insurance	not taking maintenance measures if a major storm is announced	duty of care
property insurance	not revealing valuables that could be stolen	non-disclosure
property insurance	undervaluing an asset	misrepresentation
automobile insurance	driving on worn tires	duty of care
automobile insurance	driving under the influence of alcohol	reckless conduct
life insurance	Participating in a hazardous activity or sport	reckless conduct

Table 1: Example of misconducts and legal means

such a correspondence between alleged misconducts and legal means.

Let us follow a standard way in the analysis of conflicts arbitrated by law, which consists of assuming that judges decide by relying on the likelihood of the behavior alleged by each party. An insurer may allege that the policyholder misbehaved and thus, on the basis of the law of contracts, that the claim should be fully or partially denied. However, the insurer's allegations must be consistent with the empirical evidence provided by the circumstances of losses, for otherwise the judge would consider them as insurer's bad faith and they would be invalidated. In other words, insurance law may allow insurers to condition insurance payments on loss circumstances by opening the door to legal indemnity cuts ("law completes contracts" as it is sometimes said), but its application is constrained by the approval of the judge on the basis of available information. This corresponds to the usual standard of proof for civil cases in the Common Law: judges are supposed to decide "on the balance of probabilities."[15]

[15] In such a setting, more serious allegations require more convincing evidence on the severity of the alleged misconduct. This is illustrated by Clarke (1997): "If the insurer defends with a contract exception, he must prove that exception. For example, a contract requirement that the insured "shall take all reasonable steps to safeguard the property insured" has been seen as an exception of negligence, so the claimant is not required to prove care as a condition precedent of cover, but it is for the insurer, if he wishes and can do so, to prove negligence by the insured. The onus on the insurer is to prove the exception on the balance of probabilities, but that onus will be heavier when the defense alleges fraud or willful misconduct, such as arson, by the insured." In other words, the more severe the policyholder's misbehavior alleged by the insurer, the more demanding the evidence that is required to sustain his

To deal with these issues in a simple way, suppose that when a claim is filed, the insurer privately observes a signal $s \in R$ defined by

$$s = x + \varepsilon,$$

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

When no audit is performed, the insurer routinely pays I to the claimant. If x has been verified through an audit, then the insurance payment depends upon x through legal means that may be invoked by the insurer. For notational simplicity we will not distinguish misconduct $b \in \{2, \ldots, n\}$ from the corresponding legal means (or broad standards) that can be invoked by insurers, although, in practice, there are many types of misconduct, while the law of insurance contracts only includes a limited number of broad standards. Insurance law specifies the insurer's leeway in the claim settlement process, i.e. to what extent a legal means allows him to cut or even cancel coverage. More precisely, in what follows, the law of insurance contracts is subsumed in the proportions of claims $y_b \in [0, 1]$ that the insurer is allowed to cut for each behavior b.[16] When the insurer is allowed to fully cancel the claim, we have $y_b = 1$. We assume $y_1 = 0$ because the insurer can cut the indemnity only by alleging that the policyholder misbehaved in some way. We also postulate that the law is constrained by a severity principle, according to which the severity of misconducts and the intensity of indemnity cuts are co-monotone, i.e., $y_b \leq y_{b+1}$ for all $b = 1, \ldots, n-1$. If the insurer is in a position to invoke legal means b under circumstances x, then he may decide to cut the indemnity by a fraction $z(x)$ lower or equal to y_b, and the insurance payment is $[1 - z(x)]I$.

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

allegation.

[16] We consider deterministic legal stipulations although, in practice, the application of insurance law to concrete cases is often a matter of appreciation which is within the expertise of courts. The principle of good care is a good illustration of the incomplete delimitation of the insurer's leeway when it comes to validating or rejecting claims. As written by Clarke (1997): "Insurers sometimes speak as if the insured has a legal duty to act as a prudent uninsured. However, the insured is obliged to take care to prevent or avoid *any* insured loss at all only if the contract says so in very clear terms... That being so, the same should also be true and, arguably, is true of any duty to minimize the extent or effects of loss that has already occurred or started to occur, i.e. to prevent more loss. A line between the two is hard to draw."

as

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

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

Given x, alleging misconduct $b_0 \in \{2, \ldots, n\}$ is said to be "credible on a balance of probabilities" if it is more likely that the policyholder had misconduct b_0 or a worse misconduct $b \in \{b_0+1, \ldots, n\}$ than a better behavior $b \in \{1, \ldots, b_0 - 1\}$, i.e., if[17]

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

or, equivalently if

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

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

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

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

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

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

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

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

Figure 2

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

If we take all these aspects together, we may describe the insured-insurer-judge interaction by the following five-stage game:

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

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

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

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

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

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

$$q(s) = \begin{cases} 1 & \text{if } c \leq E[y_{\hat{b}(x)}|s]I, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

It is easy to show that

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

Intuition is straightforward: claims should be audited when the signal s is bad enough to be considered as a red flag as it indicates 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 (respect. always) audited because c is very large (respect. very low).

Under a competitive insurance market, the policyholder obtains the whole surplus of her relationship with the insurer. Hence, for an optimal insurance law, the insurance contract $\{P, I, z(\cdot)\}$, the strategy $q(\cdot)$, and the insurance law itself $\{y_b, b \in \mathcal{B}\}$ maximize $\mathbb{E}u^\star$ subject to

$$P \geq \int_\Theta \pi_{b^*(\theta)} \left\{ I + \int_0^1 [c - Iz(x)]E[q(s)|x]g_{b^*(\theta)}(x)dx \right\} dH(\theta)$$

$$0 \leq I \leq L,$$

$$u^*(\theta) - d^*(\theta) \geq u_b^* - d_b(\theta) \text{ for all } b \in \mathcal{B},$$

$$q(s) = \begin{cases} 1 & \text{if } s < s^*, \\ 0 & \text{if } s \geq s^*, \end{cases} \text{, with } c = E[y_{\hat{b}(x)}|s^*]I$$

$$z(x) = y_{\hat{b}(x)} \text{ for all } x \in [0,1],$$

where the insurer's break-even constraint includes the expected audit cost c.

The optimal contract P, I under moral hazard without audit may be considered as a benchmark. It is the optimal solution of the previous problem with the restriction $q(s) = 0$ for all s, or $s^* = -\infty$. This optimal contract induces a behavior rule $b_{na}^*(\theta)$ - where the subscript na refers to "no audit" - and a threshold x_{2na} defined by

$$\sum\nolimits_{b=2}^n \Pr(b|x_{2na}) = \frac{1}{2},$$

when the policyholder's behavior is $b_{na}^*(\theta)$, with probability distribution $dH(\theta)$. In other words, x_{2na} is the threshold of accident circumstances x above which some misconduct $b \in \{2, ..., n\}$ is more probable than $b = 1$. We obtain the following result

Proposition 3 *If $\pi_b G_b(x_{2na}) > \pi_{b+1} G_{b+1}(x_{2na})$ for all $b \in \mathcal{B}$ and c is not prohibitively large, then an optimal insurance law is such that $y_b > 0$ in a non-empty subset of $\mathcal{B}$ and the insurer audits claims with positive probability (i.e., $s^* < +\infty$).*

The condition in Proposition 3 is sufficient, not necessary. It conveys the assumption that the circumstances x are strongly informative concerning the policyholder's behavior: an increase in b induces a strong FOSD shift for the probability distribution of the signal, i.e., $G_{b+1}(x_{2na})/G_b(x_{2na}) \leq \pi_b/\pi_{b+1}$.[18] In this case, under the optimal no-audit contract, the insurer infers that the policyholder

[18] $G_{b+1}(x) \leq G_b(x)$ for all x since $MLRP$ implies $FOSD$. The condition in Proposition 3 is stronger, but it is a local one at $x = x_{2na}$.

misbehaved in some way (i.e. $b \geq 2$) when x is larger than x_{2na}, and insurance law should allow him to cut indemnities in such unfavorable circumstances. This leeway provided by the law leads the insurer to condition the payment on the circumstances of the loss, which is a desirable feature of the insurance contract. This will be sustained by an equilibrium strategy where claims are verified with positive probability (i.e. when s is large enough) if the audit cost c is not too large.

Note however that insurance law limits the indemnity cuts that the insurer is allowed to decide when we have $0 < y_b < 1$. In other words, when his allegation is not dismissed, the insurer may not be allowed to fully cancel the claim. In this sense, insurance law provides leeway to reduce indemnities when some misconduct is credibly alleged, but it does not give carte blanche to the insurer. Bad signals may also be emitted by good faith policyholders (i.e. who exerted a high level of effort, e.g. who chose $b = 1$) and this is why insurance law should also limit the insurers' opportunism by preventing them from exaggeratedly reducing coverage.

5 Conclusion

The standard approach to optimal insurance under moral hazard states that insurers should incentivize policyholders by partially covering their losses. However, when contracts are complete and the circumstances of an accident are informative concerning the policyholder's effort, the indemnity should depend at the same time on the financial value of the loss, on its circumstances and on the situation – such as cases of force majeure – that may have affected the behavior of the policyholder. The less favorable the circumstances of the loss, the lower the coverage, possibly with full or zero coverage under, respectively, the most favorable or unfavorable circumstances.

When the insurance policy does not specify the indemnity payment according to all the contingencies that may characterize the claim, insurance contracts are incomplete. Conditioning the indemnity on the circumstances of the loss is nevertheless desirable because it incentivizes the policyholders to exert a high level of effort. This conditioning can be indirectly reached through legal disputes. Legal principles that allow insurers to reduce compensation under unfavorable claim circumstances act as an incentive device, but their application is constrained by the judgment of courts who decide on the balance of probabilities, and possibly by upper limits on acceptable indemnity cuts. Hence, framing the ability of the insurer to cut indemnities in the name of legal principles improves the efficiency of insurance contracting under moral hazard. We derived our results just by assuming that courts decide on the balance of probabilities. More realistic ways to model the bargain between insurer, insured and courts could be envisioned, 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. 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.[19] In practice, such a clearcut 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.[20]

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

[20] As an illustration, see Tennyson & Warfel (2009) and Asmat & Tennyson (2015) on the effect of the insurance bad faith legal regime on claims settlement, the settlement process, and insurance markets.

References

Aghion, P., Dewatripont, M., & Rey, P. (1994). Renegotiation design with unverifiable information. *Econometrica: Journal of the Econometric Society*, (pp. 257–282).

Asmat, D. & Tennyson, S. (2015). 12. the law and economics of insurance bad faith liability. *Research Handbook on the Economics of Insurance Law*, (pp. 413).

Ayres, I. & Gertner, R. (1989). Filling gaps in incomplete contracts: An economic theory of default rules. *The Yale Law Journal*, 99(1), 87–130.

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.

Edlin, A. S. (1996). Cadillac contracts and up-front payments: Efficient investment under expectation damages. *The Journal of Law, Economics, and Organization*, 12(1), 98–118.

Edlin, A. S. & Hermalin, B. E. (2000). Contract renegotiation and options in agency problems. *Journal of Law, Economics, and Organization*, 16(2), 395–423.

Edlin, A. S. & Reichelstein, S. (1996). Hold-ups, standard breach remedies, and optimal investment,(1996). *Am. Econ. Rev.*, 36, 478–478.

Grossman, S. J. & Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4), 691–719.

Hart, O. & Moore, J. (1990). Property rights and the nature of the firm. *Journal of political economy*, 98(6), 1119–1158.

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

Nöldeke, G. & Schmidt, K. M. (1995). Option contracts and renegotiation: a solution to the hold-up problem. *The RAND Journal of Economics*, (pp. 163–179).

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

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

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

Appendix

A Proof of Proposition 1

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

$$u'(W(\theta,x))\left[1 - \sum_{b \neq b^\star(\theta)} \frac{\mu_b(\theta)}{h(\theta)} \left(\frac{\pi_b}{\pi_{b^\star(\theta)}} \frac{g_b(x)}{g_{b^\star(\theta)}(x)} - 1 \right)\right] \begin{cases} \leq \lambda & \text{if } \ I(\theta,x) = 0, \\ = \lambda & \text{if } \ 0 < I(\theta,x) < L, \\ \geq \lambda & \text{if } \ I(\theta,x) = L, \end{cases} \tag{7}$$

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

Note that, for all $\theta \in \Theta$, the optimal solution is such that $b > b^\star(\theta)$ for all $b \neq b^\star(\theta)$ such that (3) is binding,[21] and also that $\mu_b(\theta) = 0$ for all $b \neq b^\star(\theta)$ such that (3) is not binding. Hence, the l.h.s. of (7) may be written as

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

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

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

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

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

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

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

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

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

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

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

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

If $b^\star(\theta) = n$, then the l.h.s. of (7) 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 Proposition 2

Consider an optimal exclusion-based optimal contract with indemnity $\overline{I}(\theta)$ and premium P. Suppose that there exists a subset $\Theta' \subset \Theta$ with positive measure, and $x_0(\theta), x_1(\theta), x_2(\theta)$ such that $x_0(\theta) < x_1(\theta) < x_2(\theta)$ and $[x_0(\theta), x_1(\theta)] \subseteq X(\theta)$ and $(x_1(\theta), x_2(\theta)] \nsubseteq X(\theta)$ for all $\theta \in \Theta'$. W.l.o.g., we may assume

$$G_{b^*(\theta)}(x_1(\theta)) - G_{b^*(\theta)}(x_0(\theta)) = G_{b^*(\theta)}(x_2(\theta)) - G_{b^*(\theta)}(x_1(\theta)), \tag{8}$$

for all $\theta \in \Theta'$. Change the indemnity schedule by substituting $[x_1(\theta), x_2(\theta)]$ to $[x_0(\theta), x_1(\theta)]$ in the exclusion area $X(\theta)$, and keep the indemnity $\overline{I}(\theta)$ and the insurance premium P unchanged. Thus $\mathbb{E}u^\star$ is unchanged and the insurer's break-even constraint still holds. Let

$$\Delta u(\theta) = u(w - P - L + I(\theta)) - u(w - P - L) > 0.$$

The incentive constraint that corresponds to behavior b in state θ, with $b > b^*(\theta)$, is written as

$$u_{b^\star(\theta)}(\theta) - u_b(\theta) \geq d_{b^\star(\theta)}(\theta) - d_b(\theta)$$

Let $\Delta_b(\theta)$ denote the change in the left-hand-side of this incentive constraint, with

$$\begin{aligned}
\Delta_b(\theta) &= \Delta u(\theta)\left(\int_{x_1}^{x_2}[\pi_b g_b(x) - \pi_{b^\star(\theta)} g_{b^\star(\theta)}(x)]dx - \int_{x_0}^{x_1}[\pi_b g_b(x) - \pi_{b^\star(\theta)} g_{b^\star(\theta)}(x)]dx\right) \\
&= \pi_b \Delta u(\theta)\left(\int_{x_1}^{x_2}\left[\prod_{i=b^*(\theta)}^{b-1}\phi_i(x) - \frac{\pi_{b^\star(\theta)}}{\pi_b}\right] g_{b^\star(\theta)}(x)dx - \int_{x_0}^{x_1}\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(x) - \frac{\pi_{b^\star(\theta)}}{\pi_b}\right] g_{b^\star(\theta)}(x)dx\right) \\
&> \pi_b \Delta u(\theta)\left[\prod_{i=b^\star(\theta)}^{b-1}\phi_i(x_1) - \frac{\pi_{b^\star(\theta)}}{\pi_b}\right]\left(\int_{x_1}^{x_2} g_{b^\star(\theta)}(x)dx - \int_{x_0}^{x_1} g_{b^\star(\theta)}(x)dx\right) \qquad (9) \\
&= 0, \qquad (10)
\end{aligned}$$

where (9) and (10) result from $\phi_b'(\cdot) > 0$ and (8) respectively. Hence the incentive constraints corresponding to behaviors $b > b^*(\theta)$ are satisfied and not-binding after the change, which contradicts the optimality of the exclusion-based contract. We deduce that any optimal exclusion-based contract is such that the exclusion area corresponds to an upper subinterval in $[0,1]$, i.e., $X(\theta) = [\widehat{x}(\theta), 1]$.

C Proof of Lemma 1

Let

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

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

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

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

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

Hence, if $b_0 \geq 2, x' > x$, we have

$$\Phi(b_0, x') > \frac{g_{b_0}(x')}{g_{b_0}(x)}\Phi(b_0, x),$$

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

D Proof of Lemma 2

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

E Proof of Proposition 3

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

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

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

$$\sum\nolimits_{b=2}^{n} \frac{\pi_b g_b(x_2)}{\pi_1 g_1(x_2)} \int_{\Theta_b^*} dH(\theta) = \int_{\Theta_1^*} dH(\theta). \tag{11}$$

The sets $\Theta_1^*, \ldots, \Theta_n^*$ depend on I, P and y, and thus (11) implicitly defines function $x_2(I, P, y)$, with $x_2(I, P, 0) = x_{2na}$ if I, P is the optimal no-audit contract.[22] Let $\lambda > 0$ and $\mu_b(\theta) \geq 0$ for $b \in \mathcal{B}, \theta \in \Theta$, be Lagrange multipliers corresponding to the insurer's break-even constraint and the incentive constraints respectively. Denote $u(1)$ and $u'(1)$ (resp. $u(2)$ and $u'(2)$ the value of the utility function and of its derivative when $x < x_2$ (resp. when $x \geq x_2$). The first-order optimality

[22] By using the MLRP assumption on $g_1(x), \ldots, g_n(x)$, we may easily show that x_2 is non-decreasing w.r.t. y. This is intuitive since an increase in the penalty rate y cannot increase the proportion of individuals who choose $b \in \{2, ..., n\}$. Thus, the balance of probabilities tips in favor of a misconduct $b \geq 2$ for more unfavorable signals.

conditions w.r.t. I and y are written as

$$\begin{aligned}
&\int_{\Theta}\left\{\pi_{b^*(\theta)}\left[u'(1)G_{b^*(\theta)}(x_2)+u'(2)(1-y)[1-G_{b^*([\theta)}(x_2)]\right.\right.\\
&\left.-\lambda\left[1-y[1-G_{b^*(\theta)}(x_2)]\right]\right\}dH(\theta)\\
&-\int_{\Theta}\{\sum\nolimits_{b\in\mathcal{B}}\mu_b(\theta)\left\{u'(1)[\pi_b G_b(x_2)-\pi_{b^*(\theta)}G_{b^*(\theta)}(x_2)]\right.\\
&\left.\left.+u'(2)(1-y)[\pi_b(1-G_b(x_2))-\pi_{b^*(\theta)}(1-G_{b^*(\theta)}(x_2))\right\}\right\}dH(\theta)\\
&-\frac{\partial x_2}{\partial I}\times A\\
&\geq 0, = 0 \text{ if } I<L,
\end{aligned} \tag{12}$$

and

$$\begin{aligned}
&I[\lambda-u'(2)]\int_{\Theta}\pi_{b^*(\theta)}[1-G_{b^*(\theta)}(x_2)]dH(\theta)\\
&+u'(2)I\int_{\Theta}\sum\nolimits_{b\in\mathcal{B}}\mu_b(\theta)\{\pi_b[1-G_b(x_2)]-\pi_{b^*(\theta)}[1-G_{b^*(\theta)}(x_2)]\}dH(\theta)\\
&-\frac{\partial x_2}{\partial y}\times A\\
&\leq 0, = 0 \text{ if } y>0,
\end{aligned} \tag{13}$$

where $A = 0$ if $y = 0$.[23] Suppose that $y = 0$ at an optimal solution, which implies $I < L$, $b^*(\theta) = b^*_{na}(\theta)$ and $x_2 = x_{2na}$. We then have $u(1) = u(2) \equiv u, u'(1) = u'(2) \equiv u'$ and (12) and (13) simplify to

$$\int_{\Theta}\{(u'-\lambda)\pi_{b^*_{na}(\theta)}-u'\sum\nolimits_{b\in\mathcal{B}}\mu_b(\theta)(\pi_b-\pi_{b^*_{na}(\theta)})\}dH(\theta)=0, \tag{14}$$

and

$$\begin{aligned}
&(\lambda-u')\int_{\Theta}\pi_{b^*_{na}(\theta)}[1-G_{b^*_{na}(\theta)}(x_{2na})]dH(\theta)\\
&+u'\int_{\Theta}\sum\nolimits_{b\in\mathcal{B}}\mu_b(\theta)\{\pi_b[1-G_b(x_{2na})]-\pi_{b^*_{na}(\theta)}[1-G_{b^*_{na}(\theta)}(x_{2na})])dH(\theta)\leq 0.
\end{aligned} \tag{15}$$

respectively. Substituting the value of $u'-\lambda$ given by (14) into (15) yields

$$\begin{aligned}
&\int_{\Theta}\sum\nolimits_{b\in\mathcal{B}}\mu_b(\theta)(\pi_b-\pi_{b^*_{na}(\theta)})dH(\theta)\times\int_{\Theta}G_{b^*_{na}(\theta)}(x_{2na})\pi_{b^*(\theta)}dH(\theta)\\
&+\int_{\Theta}\pi_{b^*(\theta)}dH(\theta)\times\int_{\Theta}\sum\nolimits_{b\in\mathcal{B}}\mu_b(\theta)[\pi_{b^*(\theta)}G_{b^*(\theta)}(x_{2na})-\pi_b G_b(x_{2na})]dH(\theta)\leq 0.
\end{aligned} \tag{16}$$

[23]The term A is not explicitly written for notational simplicity.

For all $b \in \mathcal{B}$ and all $\theta \in \Theta$, we have $b > b^*_{na}(\theta)$ and $\pi_b \geq \pi_{b^*_{na}(\theta)}$ if $\mu_b(\theta) > 0$, which implies that the first product in (16) is non-negative. Similarly, using $\pi_b G_b(x_{2na}) \geq \pi_{b+1} G_{b+1}(x_{2na})$ for all b gives $\pi_b G_b(x_{2na}) < \pi_{b^*_{na}(\theta)} G_{b^*_{na}(\theta)}(x_{2na})$ if $\mu_b(\theta) > 0$. Since this is true for all $\theta \in \Theta$, we may conclude that the second product in (16) is positive, hence a contradiction.

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

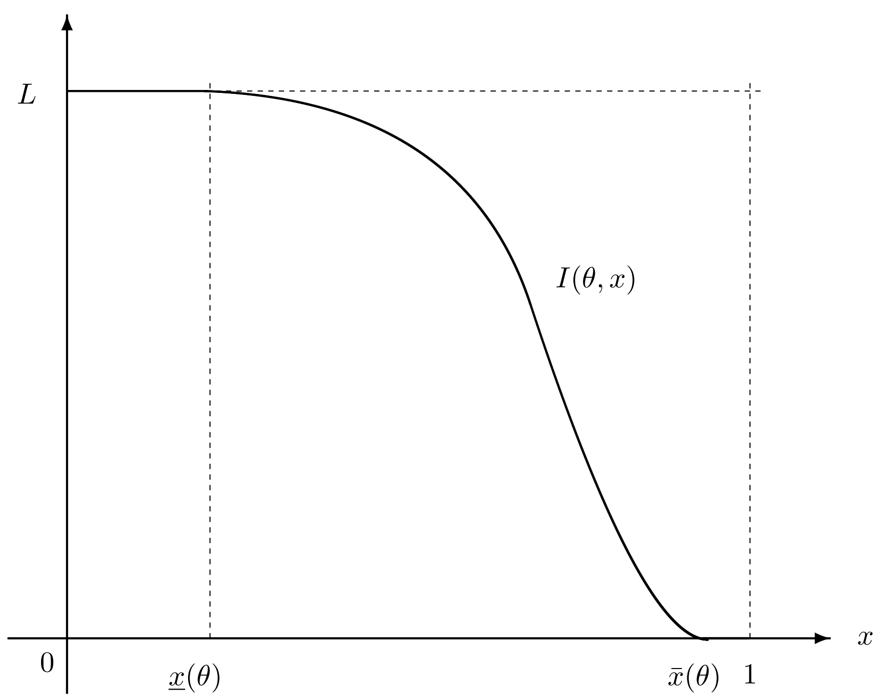

Figure 1: Indemnity schedule

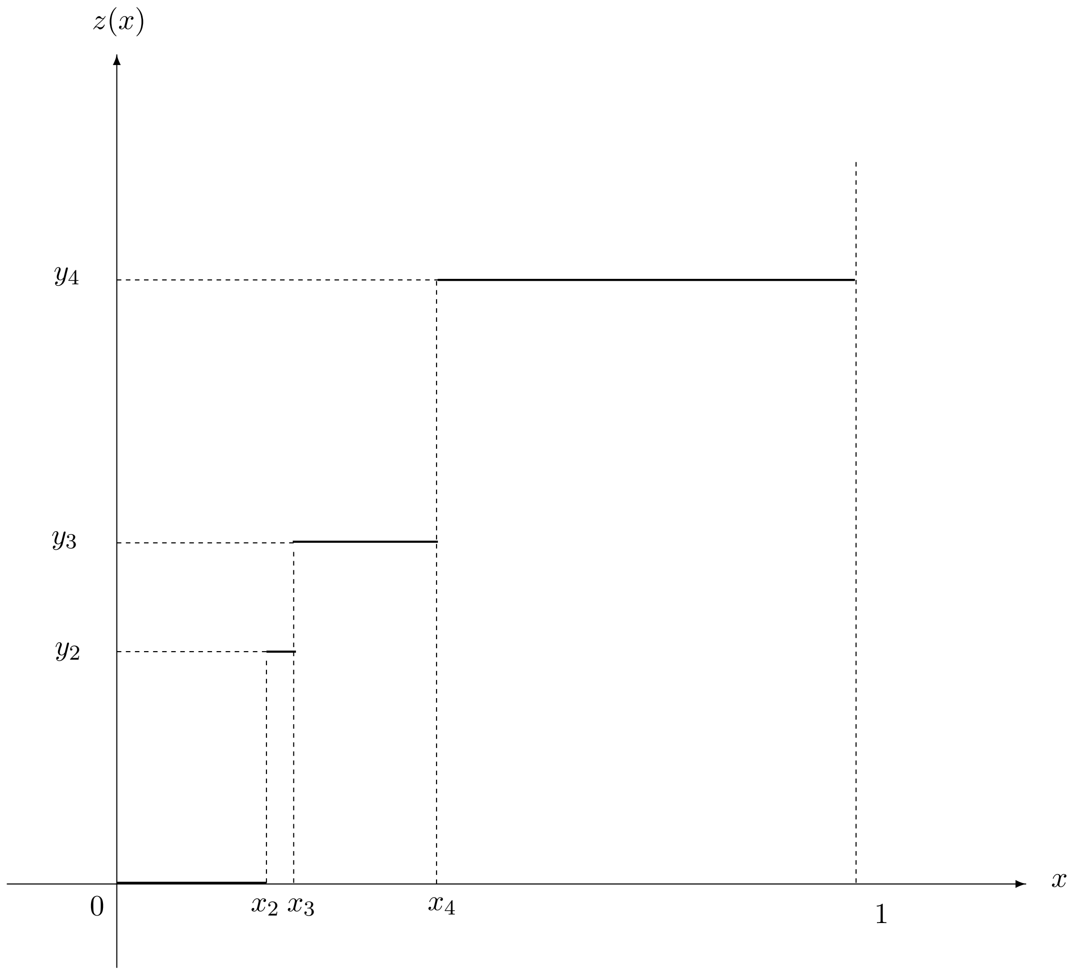

Figure 2: Optimal indemnity and the balance of probability.