Collusion in Vertical Relationships: The Case of Insurance Fraud in Taiwan

Pierre Picard, Kili C Wang

ECOLE POLYTECHNIQUE
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Collusion in Vertical Relationships: The Case of Insurance Fraud in Taiwan

Pierre PICARD
Kili C. WANG

October 4, 2016

Cahier n° 2016-16

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.portail.polytechnique.edu/economie/fr
mariame.seydi@polytechnique.edu

Collusion in Vertical Relationships: The Case of Insurance Fraud in Taiwan

Pierre Picard* Kili C. Wang†

October 4th, 2016

Abstract

The delegation of services from producers to retailers is frequently at the origin of transaction costs, associated with the discretion in the way retailers do their job. This is particularly the case when retailers and customers collude to exploit loopholes in the contracts between producers and customers. In this paper, we analyze how insurance distribution channels may affect such misbehaviors, when car repairers are joining policyholders to defraud insurers. We focus attention on the Taiwan automobile insurance market by using a database provided by the largest Taiwanese automobile insurer. The theoretical underpinning of our analysis is provided by a model of claims fraud with collusion and audit. Our econometric analysis confirms that fraud occurs through the postponing of claims to the end of the policy year, possibly by filing on single claim for several events. It highlights the role of car dealer owned insurance agents in the collusive fraud mechanism.

*Department of Economics, Ecole Polytechnique. Email: pierre.picard@polytechnique.edu. Pierre Picard acknowledges financial support from Investissements d'Avenir (ANR-121-IDEX-0003/Labex ECODEC).

†Department of Insurance, Tamkang University, Taiwan. Email: kili@mail.tku.edu.tw.

1 Introduction

Vertical relationships frequently involve the delegation of services from upstream firms to downstream retailers. This may be at the origin of transaction costs, associated with the discretion in the way retailers do their job. Such transaction costs sometimes go through the collusion between retailers and customers who exploit loopholes in the contracts between producers and customers. Discount fraud and warranty fraud are instances of such customer misbehaviors that involve collusion with retailers or frontline employees. Discount fraud exploits the special discounts that companies may offer under particular circumstances, for instance when discounted products are used for a specific purpose, e.g., educational use for softwares. Warranty fraud occurs especially when a service provider, e.g., a car repairer, replaces a defective part with a new spare part and triggers the producer's warranty, although the defective part was not original and thus was not protected by the warranty.[1]

This paper investigates another form of customer misbehavior facilitated by service providers acting on behalf of distributors: insurance fraud. Our empirical focus is on the Taiwan automobile insurance market and on the role of car dealer-owned insurance agents (DOAs) in this market. DOAs sell not only cars, but also automobile insurance to their clients, and furthermore they own an auto repair shop. Understandably, the multi-faceted activity of DOAs and their long-term connection with car owners favor the creation of a mutually advantageous policyholder-DOA alliance. Concerning fraud itself, we will focus attention on the behavior that consists in postponing claims to the last month of the policy year and in merging two losses in a unique claim. Deductibles and the bonus-malus mechanism are the underlying reasons of such behavior.

[1] See Harris and Daunt (2013) on managerial strategies under the risk of customer misbehavior. Murthy and Djamaludin (2002) survey the literature on new product warranty. Insufficient maintenance effort by buyers and inadequate behavior of retailers are at the origin of a double moral hazard problem in warranty management.

An insurance market model yields the theoretical underpinnings of our analysis. The model focuses on the strategic interaction between, on one side, policyholders who file fraudulent claims by colluding with car repairers, and, on the other side, insurers who audit claims. Auditing claims is all the more costly when collusion between policyholders and car repairers is more difficult to detect, which is particularly the case when car repairers are sheltered by DOAs. In addition, should irregularities be detected by the insurer, the high bargaining power of DOAs may allow them to deter insurers from enforcing penalties. The outcome is a higher fraud rate when insurance is distributed by DOAs than through other channels. This is reinforced in the case of deductible contracts, because deductibles weaken the insurers' incentives to monitor claims.

Our empirical analysis builds on a database obtained from the largest insurance company in Taiwan. This data includes all of the policyholders who have filed an automobile claim in 2010, amounting to nearly 11,000 files. Our results sustain the prediction that fraud is greater when insurance policies have been sold through DOAs than through other distribution channels, and also that deductibles stimulate fraud.[2] We also show that the causal mechanisms at the origin of fraud (i.e., postponing claims, and possibly filing one claim for several accidents) are linked with the bonus-malus system in force in Taiwan and with incentives that are inherent in the design of deductible contracts. This will go through an approach which consists of providing indirect evidence of such misbehaviors and of its mechanisms.[3] More explicitly, we show that the intertemporal pattern of claims

[2] Other authors have emphasized the effect of deductibles on insurance fraud. Using data from Québec, Dionne and Gagné (2001) show that the amount of the deductible is a significant determinant of the reported loss when no other vehicle is involved in the accident which led to the claim, and thus when the presence of witnesses is less likely. Miyazaki (2009) highlights through an experimental study that higher deductibles result in a weaker perception that claim padding is an unethical behavior, and thus in a larger propensity toward fraud.

[3] Although Dionne et al. (2009a) is an exception, it is usually very difficult to use direct information on fraudulent claim to analyze insurance fraud, either because identified fraud is just the top of the iceberg, or because of insurers'reluctance to share confidential information on the fraud they are victims of. The preferred approach consists of establishing indirect evidence of fraud. For instance, Dionne and Gagné (2002) and Dionne and Wang (2013) deduce the existence of fraud in automobile theft insurance from the time pattern of claims among the twelve policy months. Pao et al. (2014) provide evidence of

is consistent with policyholder's fraudulent behavior favored by DOAs, after controlling for other explanations, including adverse selection, moral hazard and the money recouping behavior highlighted by Li et al. (2013).

The paper is organized as follows. Section 2 provides further motivation for our analysis. We introduce some factual observations that should convince the reader that there is a significant degree of claim manipulation in the Taiwanese car insurance market, and we describe regular fraud patterns. Section 3 presents a model of insurance fraud that shows how these insurance fraud patterns are linked with specific features of insurance contracts, particularly deductibles, and how fraud may be facilitated by the distribution channel. Section 4 describes the data more explicitly. Section 5 presents our econometric approach and discusses our results. Section 6 concludes. The paper is completed by an appendix available online. It contains an extended version of the insurance fraud model presented in Section 3 and complementary developments of our empirical analysis.

2 Motivation

DOAs hold a substantial market share in the Taiwan automobile insurance market. For the insurance company that provides the base of our empirical analysis, 50.78% of vehicle damage insurance is sold through DOAs.[4] Furthermore, DOAs own the list of their customers, which increases their bargaining power when they negotiate contractual deals with insurance companies or when insurers monitor claims. An insurer who discovers a claim manipulation by a DOA may indeed hesitate to take retaliatory actions because of this strategic advantage of DOAs, who can choose to switch to another insurer.[5] In

opportunistic theft insurance fraud by analysing the claim pattern in areas hit by a typhoon.

[4]More precisely, 67.52% of type A contracts, 84,19% of type B contracts, and 43.71% of type C contracts are sold by DOAs. Read further for additional information on the three types of insurance contracts in Taiwan.

[5]On average, DOAs sell more policies than other agents (three times more on average and considerably more for the largest DOAs), and their market power is particularly significant for deductible contracts. They are independent agents, and, as emphasized by Mayers and Smith (1981), this status gives them

addition, DOAs also act as car repairers, and this position provides them with an informational advantage: establishing that a claim has been falsified is particularly difficult and costly when it has been filed through a DOA.

Our study is also related to the specific forms of automobile insurance fraud in Taiwan. Li et al. (2013) have observed that a large proportion of automobile insurance claims are filed during the last months of the policy year. This is confirmed by our own database. Figure 1 presents the percentage distribution of claims and the average cost of claims (in hundred US dollars) over the twelve policy months. The heavy concentration of claims in the last months of the policy year is striking. Policy years and calendar years do not coincide and, as shown in Figures 2 and 3, the concentration of claims during the last months of the policy year is compatible with seasonal fluctuations in the number of claims over the calendar year, with peaks during vacation months (January, June, July and December). In addition, the average claim amount slightly decreases in the final policy months. Li et al. (2013) interpret this phenomena as a "premium recouping effect": some policyholders without accident during the previous months tend to file false smaller claims during the last month of the policy year to express their feeling that they have been unfairly treated by the insurance company.

(Insert Figures 1, 2 and 3 here)

There are three different types of automobile physical damage insurance contracts in Taiwan: types A, B and C. Type A and B contracts cover all kinds of collision and non-collision losses, with more exclusions for B than for A,[6] while type-C contracts only cover

more discretion in claim administration (e.g., they may intercede on behalf of their customers at the claim settlement stage) because they can credibly threaten to switch their business from one insurer to another. Actually, DOAs provide comparatively less stable customers to insurers than other insurance agents, with continuation rates (i.e., the fraction of customers who keep purchasing insurance from the same insurer one year after the other) which are about sixty percent for DOAs and seventy to eighty percent for other insurance agents.

[6]Type B contracts cover all the areas of type-A contracts, except the non-collision losses caused by intentional damage, vandalism, and any unidentified reasons.

the damages incurred in a collision involving two or more vehicles. Contracts also differ in terms of indemnity: Type A contracts offer low coverage with a deductible, type B contracts may be purchased with or without deductible, and finally type C contracts provide full coverage without deductible. Claims are per accident, with a specific deductible for each claim. The change in premium is ruled by a bonus-malus system when policyholders renew their contracts with the same insurance company, with a no-claim discount and an increase in premium proportional to the number of claims, without reference to their severity. The policyholders who switch to another insurance company bargain with this company about the new starting point of the bonus-malus record

In this setting, opportunist policyholders may take advantage of manipulating claims for several reasons. According to the premium recouping interpretation of Li et al. (2013), defrauders are more likely to be among the policyholders who do not plan to keep a long term relationship with the same insurance company if, on average, such policyholders feel a lower moral cost of defrauding.[7] In our empirical analysis, this will lead us to define a "recoup group" RG that includes the policyholders who have not renewed their contract more than one year after the policy year under consideration.[8]

The bonus-malus system and the per-claim deductibles also provide incentives to defraud. Firstly, the claims filed during the last month of policy year t are not registered in the bonus-malus record of year $t+1$ (they will be taken into account in the premium paid in year $t+2$), and consequently, the policyholders who plan to renew their contract with the same insurer may be incited to postpone their claim to the last policy month, in order to delay the increase in premium.[9] Secondly, since the bonus-malus record depends

[7] It is well known that insurance fraud is often associated with the feeling that the insurance company is unfair; see Fukukawa et al. (2007), Miyazaki (2009) and Tennyson (1997, 2002). The premium recouping phenomenon could reflect a kind of resentment against insurers, particularly from policyholders who have not filed any claim during the policy year.

[8] Because of the bonus-malus system (see below), the policyholders who renew their contract only one year have the same incentive to defraud as the policyholders who switch insurers at the end of the policy year.

[9] In addition, the bonus-malus system forgives the first accident for drivers who have had no other

on the number of claims and not on their severity, policyholders may be prompted to file one unique claim for two accidents, should a second accident occur. This is even more profitable in the case of deductible contracts, since deductibles are per-claim: the strategy that consists of postponing the first claim and merging any other accident with the first one within a unique claim yields full coverage for the part of the year that follows the first accident. Type A and B contracts are subject to this kind of claims manipulation, because they include coverage for losses other than those associated with the collision between two cars. There is no third-party involved in such claims and no police report. On the contrary, the claims filed for type C contracts correspond only to collisions, and they have to include a police report, which makes manipulation very unlikely. In our empirical analysis, the set of policyholders who renew type A or B contracts with the same insurer will be called the "suspicious group" SG because of this incentive to manipulate the bonus-malus system, with subgroups $SG1$ and $SG2$ for no-deductible and deductible contracts, respectively.

If we conjecture that some claims filed in the last policy month correspond in fact to postponed claims with the cumulated losses of two events, then we should expect that the ratio of "the average cost of first claims" over "the average cost of all claims" (hereafter called the *first claim cost ratio*) should increase during this month. Note however that this prediction could also be interpreted as the outcome of a moral hazard mechanism: this would be the case if a first accident made drivers more cautious, and thus they have less severe accidents should a second accident occur during the same policy year. To disentangle these two explanations, we may consider type C contracts as a benchmark to isolate the moral hazard effect, since claims manipulation is unlikely for such a contract.[10] Figure 4 confirms our intuition: the first claim cost ratios for $SG1$ and $SG2$ significantly

accidents for three years, which provides an even larger manipulation gain.

[10]Type C contracts only cover the risk of collision. Thus, their claims involve a third party, which makes manipulation difficult.

jump in the last month, and this is not the case for type C policies.

(Insert Figure 4 here)

At this stage, we may come back to the part played by DOAs. Figure 5 confirms that DOAs may favor the manipulation of claims. While the claims filed by the policyholders of the two suspicious groups, $SG1$ and $SG2$, are significantly concentrated in the last policy month, this pattern is even more obvious for the policyholders of each subgroup that have purchased insurance from DOAs. Figure 5 also shows that the last policy month pattern is much less obvious in the benchmark group, which includes those policyholders who are covered by no-deductible contracts and who have not renewed their contract with the same insurance company at the end of the policy year.

(Insert Figure 5 here)

3 The model

The core of the following model is the strategic interaction between policyholders who defraud by colluding with car repairers, and insurers who allocate resources to monitor claims. The objective of this model is to highlight the logical link between, on one side, the intensity of fraud and, on the other side, relevant features of insurance contracts, particularly the size of deductibles and the distribution channel.[11]

Consider a population of risk-averse drivers, whose type is defined by the couple $(i, h) \in \{D, A\} \times \{1, 2\}$. Index i refers to the individuals' preference for a specific distribution channel through which they purchase insurance: DOA when $i = D$ or standard insurance

[11] The model features the non-cooperative interaction between policyholders and insurers, in a costly state verification setting as in Picard (1996). For the sake of brevity, several aspects of the insurance market analysis are deliberately overlooked here. This concerns particularly the way individuals choose their contract and their insurance distribution channel, depending on their risk aversion and on their intinsic preference for a specific channel. A more complete version of the model is in Section 1 of the Appendix (available online).

agents when $i = A$. Index h reflects the individual's degree of absolute risk aversion: $h = 1$ corresponds to a higher absolute risk aversion than $h = 2$. Drivers may have either 0,1 or 2 accidents during the same policy year, with probability π_1 and π_2 for 1 and 2 accidents, respectively, and $\pi_1 + \pi_2 < 1$. These probabilities are independent from the policyholders' type. Insurance contracts include a deductible per accident. We respectively denote d_{ih} and P_{ih} the deductible and the premium of the contract chosen by type h individuals who purchase insurance through channel i. Less risk averse individuals choose a larger deductible, and thus we have $d_{i2} > d_{i1} \geq 0$.

Each accident may be severe or minor, with probability q_s or $q_m = 1 - q_s$, respectively (independently from the policyholder's type) and the corresponding claims are small or large, respectively. To simplify matters, its is assumed that a large claim exactly doubles a small claim. Fraud is committed by policyholders who postpone small claims till their last policy month. They will file one single large claim for two minor accidents presented as a severe accident, should another minor accident occur later during the same policy year. Otherwise, their claim will be denied. Fraud reduces the retained cost of the acccidents by half since the deductible is paid only once. It also provides a supplementary gain through the manipulation of the bonus-malus system if the policyholder intents to stay with the same insurer at least during the next year.

Defrauding requires colluding with a car repairer, and, in the case of successful fraud, the policyholder and the repairer will share these benefits according to their respective bargaining powers. The audit of claims by the insurer makes fraud risky: each member of a policyholder-repairer coalition that is spotted defrauding has to pay a penalty (considered, for simplicity, as a fine to the government), and the claim is fully denied (i.e., the policyholder does not receive any indemnity).

Let us denote by α_{ih} and β_{ih} the fraud and audit mixed strategy of the policyholder and of the insurer, respectively, for a population of type (i, h) individuals. α_{ih} is the probability that a type (i, h) policyholder postpones a first claim (when the corresponding minor accident occurs before the last policy month), with the intention to file a unique large claim for two accidents, should another minor accident occur before the end of the policy year. Fraud is concentrated among the policyholders who are willing to stay with the same insurer, because they are the ones who benefit most from fraud through the bonus-malus mechanism.[12] β_{ih} is the probability that a large claim (filed by a type (i, h) policyholder) is audited by the insurer.[13] Such large claims correspond either to true severe accidents or to two minor accidents that have been fraudulently aggregated). We assume that audit allows the insurer to detect with certainty whether the claim is fraudulent or not, i.e., whether it corresponds to two small claims that have been falsified as a single large claim, or whether it corresponds to a true large loss.

The expected cost of claims per type (i, h) policyholder is written as

$$C_{ih} = \overline{L} - (\pi_1 + 2\pi_2)d_{ih} + FC_{ih} + AC_{ih}, \tag{1}$$

where $\overline{L}$ is the expected costs of accidents (depending on the cost and probability of minor and severe accidents), $(\pi_1 + 2\pi_2)d_{ih}$ is the cost retained by the policyholder (in the absence of fraud), FC_{ih} is the cost of fraudulent claims and AC_{ih} is the audit cost. FC_{ih} is proportional to α_{ih}, but, for given α_{ih}, it decreases linearly with β_{ih}, because auditing a larger fraction of large claims reduces average indemnity payment after detecting falsified claims, i.e., claims that result from the merging of two small claims. DOAs have some bargaining power with insurers and they may intercede with the insurer when

[12] The other policyholders will not have enough incentives to defraud.

[13] Note that the degree of risk aversion is not directly observed by the insurer. However, individuals choose different contracts (i.e., different deductibles) depending on their risk aversion, and thus insurers can condition their audit probability on the level of the deductible, and thus indirectly on the policyholder's type. Note also, that the policy year and the calendar year do not coincide. The beginning of the calendar year is evenly distributed over the calendar year among the policyholders. Only the first claims that correspond to (true or falsified) severe accidents are audited. For practical reasons, it is assumed that insurers audit all these claims with the same probability, whether they are filed within or outside the last month of the policy year.

a claim is denied for fraud. This intervention is successful with some probability, and then it decreases the financial benefit drawn by the insurer from spotting a defrauding policyholder-car repairer coalition. Thus, we may write

$$FC_{ih} = \alpha_{ih}[a_1(d_{ih}) - a_2(d_{ih}, \zeta_i)\beta_{ih}], \tag{2}$$

with $a_1(d_{ih}), a_2(d_{ih}, \zeta_i) > 0$, where ζ_i is a parameter that measures the bargaining power of distribution channel i, with $\zeta_D > \zeta_A$.[14] We have $a'_1 > 0$ and $a'_{2d} < 0$ because the larger the deductible, the larger the financial impact of claims falsification and the smaller the gain to the insurer when a claim is denied after audit. We also have $a'_{2\zeta} < 0$ because the distribution channel's bargaining power leads to a smaller insurer's expected benefit when fraud is detected.

DOAs own and control their repair shop. Thus, it is assumed that auditing a claim (i.e., spending resources to discover whether a claim has been falsified or not) is more costly when insurance has been purchased through a DOA than through a standard insurance agent, because the protection of the DOA makes the detection of collusion more difficult. Let c_i be the audit cost when the insurance distribution channel is $i = D$ or A, with $c_D > c_A$. The number of large claim filed by type (i, h) policyholders is linearly increasing with α_{ih}, which allows us to write[15]

$$AC_{ih} = c_i\beta_{ih}(a_3 + a_4\alpha_{ih}). \tag{3}$$

[14] Fraud, as it is described, may be committed by policyholders who intend to renew their insurance policy and who have two small accidents, the first one being severe and occurring before the last month of the policy month. Thus, $a_1(d_h)$ and $a_{2i}(d_h)$ depend on the probability that a type (i, h) individual is in this situation and it depends upon π_1, π_2 and q_s. See Section 1 of the Appendix for a more explicit formulation.

[15] a_3 and a_4 depend upon π_1, π_2 and q_s. See Section 1 of the Appendix for details.

The insurer chooses β_{ih} in $[0,1]$ in order to minimize C_{ih}, which gives

$$\beta_{ih}\begin{cases} = 0 \text{ if } \alpha_{ih} < \alpha^*(d_{ih}, \zeta_i, c_i), \\ \in [0,1] \text{ if } \alpha_{ih} = \alpha^*(d_{ih}, \zeta_i, c_i), \\ = 1 \text{ if } \alpha_{ih} > \alpha^*(d_{ih}, \zeta_i, c_i), \end{cases} \tag{4}$$

where

$$\alpha^*(d, \zeta, c) \equiv \frac{ca_3}{a_2(d, \zeta) - ca_4}. \tag{5}$$

with $\alpha_d^{*\prime} > 0, \alpha_\zeta^{*\prime} > 0$ and $\alpha_c^{*\prime} > 0$. Let us assume for simplicity that $\alpha^*(d, \zeta, c) < 1$ for the relevant values of d, ζ, c. This means that systematic fraud would trigger the auditing of all the large claims. Depending on the bribe that they have to pay to car repairers for them to collude (which is not explicitly defined here), policyholders are willing to defraud if the probability of being caught is larger than a threshold $\beta_h^*(P_{ih}, d_{ih}) \in (0, 1)$. Individuals always defraud when the audit probability is zero, and they never defraud if all large claims are audited: hence the audit probability $\beta_h^*(P_{ih}, d_{ih})$ for which they are indifferent between fraud and honesty is between 0 and 1. This audit probability threshold is type dependent (hence the subscript h in the β_h^* function) because it is affected by the intrinsic risk aversion of the policyholder, but it also depends on P_{ih} because an increase in premium may affect the policyholder's risk aversion through a wealth effect,[16] and it is increasing with d_{ih} because an increase in the deductible makes fraud more attractive.

[16] For instance, under DARA, an increase in the insurance premium makes the policyholder more risk averse, and thus less prone to conclude a risky fraudulent arrangement with a car repairer. In that case, the larger the insurance premium, the lower the audit probability threshold above which fraud is detered.

Thus, we have

$$\alpha_{ih} \begin{cases} = 0 \text{ if } \beta_{ih} > \beta_h^*(P_{ih}, d_{ih}), \\ \in [0,1] \text{ if } \beta_{ih} = \beta_h^*(P_{ih}, d_{ih}), \\ = 1 \text{ if } \beta_{ih} < \beta_h^*(P_{ih}, d_{ih}). \end{cases} \tag{6}$$

A type (i, h) policyholder who has a minor accident before the last policy month and her insurer play a non-cooperative game, with strategies α_{ih} and β_{ih} respectively. Its Nash equilibrium is easily characterized. If $\alpha_{ih} = 0$, then (4) gives $\beta_{ih} = 0$, which implies $\alpha_{ih} = 1$ from (6), hence a contradiction. Similarly, if $\alpha_{ih} = 1$, then (4) gives $\beta_{ih} = 1$, which implies $\alpha_{ih} = 0$ from (6), hence again a contradiction. Thus, $\alpha_{ih} \in (0, 1)$ and (4),(6) give $\beta_{ih} = \beta_h^*(P_{ih}, d_{ih}) \in (0, 1)$ and $\alpha_{ih} = \alpha^*(d_{ih}, \zeta_i, c_i)$.

Thus, at equilibrium, the audit probability $\beta_{ih} = \beta_h^*(P_{ih}, d_{ih})$ makes the policyholder indifferent between fraud and honesty, and the fraud probability $\alpha_{ih} = \alpha^*(d_{ih}, \zeta_i, c_i)$ makes the insurer indifferent between auditing and not-auditing.

This leads us to simple predictions about the effect of the type of contract and distribution channel on insurance fraud. Firstly, using $\alpha_d^{*\prime} > 0$ shows that higher deductibles go along with more fraud. Since $d_2 > d_1 \geq 0$, we have $\alpha_{i2} > \alpha_{i1}$: in other words, for a given distribution channel, fraud is more prevalent among type 2 than type 1 individuals. More simply, if $d_1 = 0$, we can say in a shortcut that deductibles encourage fraud. We also have $c_D > c_A$ and $\xi_D > \xi_A$, and thus using $\alpha_\zeta^{*\prime} > 0$ and $\alpha_c^{*\prime} > 0$ yields $\alpha_{Dh} > \alpha_{Ah}$. Put briefly, for a given type of individuals, there is more fraud when insurance has been purchased through the DOA agents than through standard insurance agents, because of the ability of DOAs to shelter their car repairers either from full-fledged inquiries or from sanctions.

4 The data

Our data comes from the largest insurance company in Taiwan, with an automobile insurance market share of over 20%. Data is recorded at the individual level and provides detailed information about the policyholders, their insurance contracts and the claims they have filed. Available variables are listed in Table 1. Data has been collected over the 2010-2012 period, but our analysis will be restricted to 2010, so that we know whether policyholders subsequently renewed their contracts for less or more than one year.[17] We target the owners of private usage small sedans and small trucks with type A, B or C insurance contracts for automobile physical damages. There are $109,461$ policyholders, and 45.86% of them filed at least one claim in the year 2010, which corresponds to $50,194$ observations. This subset defines our "research sample" (i.e., the sub-sample of policyholders with claims).

(Insert Tables 1, 2-1 and 2-2 here)

The mean values of the variables in the two samples are displayed in Table 2-1, with some significant differences. In particular, the research sample includes a larger proportion of female, middle-aged owners, large-sized and new vehicles. The insured in the research sample also tend to purchase higher coverage contracts than those in the whole sample. Both the percentages of type A and B contracts are comparatively higher in the research sample (1.52% vs 1.03%, 67.42% vs 38.82%). The vehicles are also more concentrated in some particular brand. More importantly, the research sample includes a larger fraction of policyholders who belong to the $SG1$ and $SG2$ group, and who have purchased insurance through the DOA channel than the whole sample (15.17% vs 11.77%, 28.15% vs 14.48%, and 62.16% vs 50.78%, respectively). The share of the RG group also increases from 19.8% in the whole sample to 28.47% in the research sample. Furthermore, the claim

[17] In what follows, years are policy years: a contract corresponds to year 2010 if it started in 2010.

record measured either by the average bonus-malus coefficient or by the average premium is worse in the research sample than in the whole sample.

Table 2-2 separates the research sample into two subgroups, according to the insurance distribution channels (DOA and non-DOA), with significant differences in terms of gender, usage, and vehicle size. There is also a higher proportion of new vehicles for the DOA channel, which reflects the fact that DOAs sell both vehicles and insurance contracts. On average, the bonus-malus coefficient is significantly higher in the DOA group than in the non-DOA group, but insurance premiums do not significantly differ between the two groups.[18] Furthermore, the percentage of insured parties who belong to the SG group is significantly different between these two channels, for SG1 (21.66% vs 4.50%, respect.) as well as $SG2$ (26.66% vs 30.59, respect.). The percentage of claims filed in the suspicious period (defined as the last month of the policy year) is 8.49% in the non-DOA channel, and it rises to 22.99% in the DOA channel. We may also observe that the share of the RG group is significantly lower in the DOA channel (13.73% vs 37.44%). Finally, the percentage of deductible contracts sold through the DOA channel is smaller than that in the non-DOA channel (54.61% vs 60.12%).

5 Testing hypotheses

5.1 Evidence on claim manipulation

We firstly test the hypothesis that the perspective of contract renewal and the choice of a deductible contract are factors that stimulate fraud. We focus on the fraudulent behavior that consists of manipulating the claim date by moving it to the last policy month, called the "suspicious period", possibly by filing one claim for two events. We define the fraud

[18] It is indeed well known in Taiwan that individuals with less favorable claim records (and thus with a higher bonus-malus coefficient) tend to purchase insurance through a DOA, and that some DOAs may unduly protect their customers from a strict enforcement of the bonus-malus rule.

rate as the number of claims per policyholder filed during the suspicious period, hence the following wording of Hypothesis 1.[19]

Hypothesis 1: *The fraud rate is higher in the suspicious group than in the non-suspicious group, and this is particularly the case for individuals covered by deductible contracts.*

Testing Hypothesis 1 amounts to identifying whether there is a conditional dependence between belonging to the suspicious group SG (or to one of its subgroups $SG1$ and $SG2$) on one side, and filing a claim within the suspicious period (evaluated by the dummy SC) on the other side. To do so, we use a two-stage method, similar to the approach followed by Dionne et al. (1997, 2001).[20] For notational simplicity, $SG, SG1, SG2$ also denote dummies for belonging to suspicious groups $SG, SG1$ and $SG2$, respectively. *deduct* is a dummy for deductible contracts. SG and *deduct* are estimated at Stage 1 by bivariate Probit regressions, with an instrumental variable approach, with $SG1 = SG \times (1 - deduct)$ and $SG2 = SG \times deduct$. Stage 1 requires finding out some factors that are related to the renewal and coverage decisions, in addition to the underwriting and pricing variables, and that are unrelated to the decision of filing a suspicious claim.

Income and education may conceivably affect the type of contract and the mobility between insurers, but there is no obvious reason for which they should be related to the decision of filing a claim during the last policy month. Thus, they are natural candidates to be instruments at stage 1. Unfortunately, our database does not provide information about the income and education levels of each policyholder. As approximations, we use *income* and *edu*, that measure the average income and the percentage of the population

[19] Of course, this does not mean that all the claims filed in the suspicious period are fraudulent.

[20] They aim at appraising whether the correlation between claims and coverage reflects individuals' unobservable characteristics, which are not used by insurers in underwriting and pricing decisions. In order to avoid a spurius correlation caused by misspecification, they add the conditional expectation of one decision variable (such as filing a claim) when regressing on the other one (such as choosing the insurance coverage). To avoid endogeneity problems, Dionne et al. (2009b) and Dionne et al. (2015) estimate this conditional expectation through an instrumental variable approach.

with a master or PhD degree, respectively, in the zip code area of the policyholder.[21]

Accordingly, at stage 1, bivariate Probit regressions are written as

$$\Pr(SG_i = 1|income_i, edu_i, X_i) = \Phi(\beta^1_{inc} income_i + \beta^1_{edu} edu_i + \alpha^1 X_i + \varepsilon_{SGi}), \quad (7)$$

$$\Pr(deduct_i = 1|income_i, edu_i, X_i) = \Phi(\beta^2_{inc} income_i + \beta^2_{edu} edu_i + \alpha^2 X_i + \varepsilon_{dedti}), \quad (8)$$

with $cov(\varepsilon_{SGi}, \varepsilon_{dedti}) = \rho$ and where X_i is the column vector of underwriting and pricing variables for policyholder i, including: gender and age of the policyholder, usage, brand, size and age of the insured vehicle, the bonus malus coefficient and the premium level. This is the first group of explanatory variables in Table 1.

At stage 2 of the 2SLS approach, we estimate the probability that policyholders file their first claim during the suspicious period. We explore the conditional dependence between SC and $SG1$ and between SC and $SG2$, by considering $\Pr(SG1_i) \equiv \Pr(SG_i = 1, deduct_i = 0)$ and $\Pr(SG2_i) \equiv \Pr(SG_i = 1, deduct_i = 1)$ as explanatory variables in a second stage Probit regression, which is written as

$$\begin{aligned} &\Pr(SC_i = 1|\Pr(SG1_i), \Pr(SG2_i), RG_i, X_i) \\ &= \Phi(\beta_{instr1} \Pr(SG1_i) + \beta_{instr2} \Pr(SG2_i) + \beta_r RG_i + \beta X_i), \end{aligned} \quad (9)$$

where $SC_i = 1$ when policyholder i has filed his first claim during the suspicious period and $SC_i = 0$ otherwise. To control for the possibility that last policy-month claims may result from a premium recouping behavior, we also use the control variable RG_i (with $RG_i = 1$ when the contract is of type A or B and has not been subsequently renewed for more than one year, and $RG_i = 0$ otherwise). Here, also, X_i is the column vector that

[21] We have also tested the percentage of population with education levels higher than the bachelor degree, or higher than the junior college. Using the percentage of inhabitants with a master or PhD degree was the best way to measure the effect of education on SG and *deduct*.

contains first group explanatory variables of Table 1.

Alternatively, as Dionne et al. (2015), we may also include dummies for the variables instrumented at stage 1 among the explanatory variables of stage 2 regression. This method (hereafter referred as the DGV approach) leads to write the stage 2 regression as

$$\begin{aligned}&\Pr(SC_i = 1|\Pr(SG1_i), \Pr(SG2_i), SG1_i, SG2_i, RG_i, X_i)\\&= \Phi(\beta_{instr1}\Pr(SG1_i) + \beta_{instr2}\Pr(SG2_i) + \beta_{s1}SG1_i + \beta_{s2}SG2_i + \beta_r RG_i + \beta X_i).\end{aligned} \tag{10}$$

In the 2SLS approach (regression (9)), the conditional dependence between $SG1$ and SC as well as between $SG2$ and SC is evaluated through the estimated coefficients of $\Pr(SG1_i)$ and $\Pr(SG2_i)$, i.e., by β_{instr1} and β_{instr2}, respectively. In the DGV approach (regression (10)), the conditional dependence is evaluated by the overall sum of the estimated coefficients of $\Pr(SG1_i)$ and $SG1_i$ and the sum of the estimated coefficients of $\Pr(SG2_i)$ and $SG2_i$, i.e., by $\beta_{instr1} + \beta_{s1}$ and $\beta_{instr2} + \beta_{s2}$, respectively.[22]

(Insert Table 3 here)

The first stage bivariate Probit estimations are listed in the two first columns of Table 3, with intuitive results. Wealthier people have a lower probability of continuing the same contract, and a higher probability of purchasing a deductible contract. This is consistent with a decreasing absolute risk aversion assumption: in a setting where individuals may have partial information on the quality of insurance contracts, less risk

[22] As a preliminary step, the 2SLS approach requires testing (1) whether there is a weak instrument problem by the Anderson-Rubin test, (2) whether the instrument is over-identified by Sargan's J test, and finally (3) whether the instrumental variable method is relevant by the Durbin-Wu-Hausman test. Dionne et al. (2015) state that estimating the conditional probability of the instrumented variable through LPM or through the Probit model is qualitatively consistent with the 2SLS approach. Estimating $\Pr(SG_i = 1|income_i, edu_i, X_i)$ and $\Pr(deduct_i = 1|income_i, edu_i, X_i)$ by two LPMs and performing these three tests validates our IV approach. The results of these tests are in Table 8 in Section 2 of the Appendix.

averse individuals are less reluctant to switch insurer, and they also tend to choose lower coverage. Furthermore, the education level is negatively correlated with the deductible dummy, that is highly educated people tend to purchase more insurance, and positively correlated with the renewal decision. Members of the recoup group renew their contract less frequently (which simply reflects the definition of RG), and they tend to opt for deductible contracts. We also see that the owners of larger vehicles are comparatively more likely to renew their insurance contract and to opt for a contract with a deductible.

The results of the second-stage estimation by the 2SLS approach are reported in the third column of Table 3. They show the conditional dependence between SC and either $SG1$ or $SG2$, with coefficients 0.6110 and 0.8021 that are significant at the 10% and 1% levels, respectively. The fourth column corresponds to the second stage of the DGV approach. The estimated coefficients of $\Pr(SG1)$ and $\Pr(SG2)$ are significantly different from 0 at the 10% and 1% levels, respectively, which confirms the existence of an endogeneity problem. The dummy variables $SG1$ and $SG2$ are also significantly different from 0 at the 10% and 1% thresholds, respectively, which confirms the conditional dependency between SC and $SG1$ or $SG2$, with total coefficients $1.8305 = 0.6094 + 1.2211$ and $2.4919 = 0.7809 + 1.7110$, respectively.

Thus, the 2SLS and DGV approaches lead to similar conclusions, and they confirm our presumption of a positive conditional dependence between belonging to $SG1$ or $SG2$ and filing a first claim during the suspicious period, which supports Hypothesis 1.[23]

[23] Table 3 also offers some interesting byproducts that are worth mentioning. Firstly, the owners of new vehicles tend to file their first claim during the suspicious period, which reflects the so-called "car wash" phenomenon in Taiwan, that is the fact that purchasers of new cars may benefit from a free visit to the car repairer at the end of the first year. As we will see, this is linked with the role of car dealers in insurance fraud. Secondly, the policyholders from the RG group also tend to file their first claims in the suspicious period, which echoes the conclusions of Li et al. (2013). Thirdly, females file their first claim during the suspicious period more frequently than males, but that does not necessarily reflect a gender effect in fraudulent behavior. It is usual in Taiwan to register cars under the name of females (e.g., a wife or mother), even when the primary driver is a male, in order to benefit from cheaper insurance premiums. Hence, instead of a gender effect, the above mentioned correlation may just reflect the fact that the policyholders who carefully manage their insurance budget may also try to obtain undue advantage from their insurance company.

If defrauders postpone their claims to the suspicious period and if they may cumulate losses in a unique claim, then the suspicious period should be characterized by high values of first-claim cost ratios. This is expressed in Hypothesis 2.

Hypothesis 2: *The first-claim cost ratio is larger in the suspicious period than during the rest of the policy year, particularly for the suspicious group.*

Hypothesis 2 is tested through the following regression:

$$clmamt_i = \alpha_c SC_i + \alpha_f first_i + \alpha_{fs} first * SC_i + \alpha X_i, \tag{11}$$

which is performed among the claims filed by members of the SG group, where $clmamt_i$ is the value (in US dollars) of the claims filed by policyholder i. In regression (11), we use two additional variables ($first_i$ and $first_i * SC_i$) besides SC_i and vector X_i. $first_i = 1$ when this is the first claim filed by policyholder i during the policy year, otherwise $first_i = 0$ and $first * SC_i$ is an interaction variable. We perform the above test separately for $SG1$ and $SG2$. In our sample, this corresponds to 22,081 claims filed by 7,614 policyholders from SG1, and 25,434 claims filed by 14,130 policyholders from SG2. The estimated coefficient of the interaction term $\widehat{\alpha}_{fs}$ is the key to test. We obtain $\widehat{\alpha}_{fs} = 108.54$ with $p-$value 0.120 for $SG1$, and $\widehat{\alpha}_{fs} = 238.09$ with $p-$value 0.035 for $SG2$. To some extent, these results confirm the validity of Hypothesis 2, with a lower significance level for $SG1$ than for $SG2$.[24] For the sake of completeness, we have also run the regression that explains the value of the claims over the whole sample (not only the SG group) by including dummies $SG1_i, SG2_i, SC_i, first_i, RG_i$ and their double and triple interaction terms in the explanatory variables. Results confirm the validity of Hypothesis 2 (see Section 2 in the Appendix).

[24] This is consistent with the fact that the policyholders with deductible contracts (i.e., the $SG2$ subgroup) have a greater incentive to file a unique claim for two events than the policyholders with no-deductible contracts (the $SG1$ subgroup).

5.2 Robustness tests

1. To check the robustness of our conclusions, we have also tested Hypothesis 1 by following an approach inspired from Chiappori and Salanié (2000). The results are reported in Section 2 of the Appendix and they confirm our conclusion about the conditional correlation between SG_1 or SG_2 and SC.

2. It is also worth investigating whether the validity of our conclusions can be affected by the hypothetical presence of (ex ante) moral hazard or adverse selection. Ex ante moral hazard explains why a more comprehensive insurance coverage may make a driver less cautious. This incentive effect is even stronger for policyholders who had no accidents before the suspicious period, because the bonus-malus system forgives the first accident. Hence, under the moral hazard hypothesis, the policyholders from the $SG1$ group (i.e., those with a no-deductible contract) should be less cautious than those from $SG2$ (the policyholders with a deductible contract), and according to the moral hazard interpretation, they should have more severe first accidents in the last policy month. The comparison of coefficients $\widehat{\alpha}_{fs}$ in regression (11) for $SG1$ and $SG2$ leads to the opposite conclusion.

3. In a setting with adverse selection, past and future claim experiences may be linked, but man-made claim manipulation should reduce the predictive power of this link. Furthermore, adverse selection may lead to a positive correlation between the coverage and the probability of filing claims, but it does not induce any particular timing for claims. These observations open the door to additional tests reported in Section 2 of the Appendix, which confirm that our conclusions on claim manipulation are not called into question by hidden information on risk types.

4. Finally, we may also be worried by the fact that the $SG2$ group includes two types of deductible contracts, with more extensive exclusions for type B than for type A.

To control for any disturbances that may be linked to this difference in the scope of coverage, we have redone our tests by limiting our sample to type-B contracts, with unchanged conclusions. Results are reported in Section 2 of the Appendix.

5.3 Evaluating the cost of fraud

Beyond the mere fact that fraud does exist, estimating its cost is also important. To do this, we refer to the empirical results from the DGV model. The estimated coefficients of $\Pr(SG1)$ and $SG1$ are 0.6094 and 1.2211 (see the fourth column in Table 3), and their marginal effects are 0.1664 and 0.3335, respectively. This implies that, overall, the probability of filing a claim in the suspicious period increases by 49.99% when comparing a policyholder from the $SG1$ group to those in the non-suspicious group. The average cost of non-detected fraudulent claims is US\$199 (NT\$$5,970$) if we presume that fraud is committed by filing a unique claim for two events, postponed to the last month of the policy year to avoid the penalty from the bonus-malus rule.[25] This implies that the difference in annual fraud cost between members of the $SG1$ group and policyholders from the non-suspicious group is about $199 \times 0.4999 =$US\$99.48. Likewise, the estimated coefficients of $\Pr(SG2)$ and $SG2$ are 0.7809 and 1.7110, with marginal effects 0.2133 and 0.4673, respectively. This implies that the probability of filing a claim in the suspicious period increases by 68.06% when we compare members of the $SG2$ group to policyholders from the non-suspicious group. The average cost of a fraudulent claim is US\$365.67, once again with the assumption that defrauders file a unique claim for two events and postpone their claim to the last month of their policy year.[26] This implies that the policyholders

[25] The average insurance premium in our research sample is NT\$ 14,925 (US\$497.5). We may roughly estimate that the defrauders who file a unique claim for two events and postpone their claim to the last month of their policy year avoid about 40% of this amount.

[26] This cost includes the avoided deductible and the avoided bonus-malus penalty. The deductibles of first and second claims are NT\$$3,000$ and NT\$$5,000$ respectively, hence there is a NT\$$5,000$ fraud cost when policyholders file a unique claim to cover two accidents. Adding the NT\$$5,970$ avoided penalty due to the increase in premium to the deductible of the second claim yields a total fraud cost of

from the $SG2$ group entail an expected cost of fraud that is about US$248.87 higher than for the insured from the non-suspicious group. Since there are $7,614$ and $14,130$ policyholders in $SG1$ and $SG2$, respectively, we may deduce that the expected cost of fraud is about US$$4,273,974$ which represents 11.54% of the total premiums paid by the policyholders from our sample (US$37 million). These are of course very crude estimates, but they give an idea of the cost of fraud through claims manipulation in Taiwan.

5.4 On the role of DOAs

Our third hypothesis relates the fraud rate to the insurance distribution channel.

Hypothesis 3: *The fraud rate in the suspicious group is comparatively even larger when insurance has been purchased through the DOA channel than through other distribution channels.*

Testing the validity of Hypothesis 3 will follow the same approach as for Hypothesis 1. Dummy D_i indicates that policyholder i has purchased insurance through the DOA channel, and now three endogenous variables, $SG_i, deduct_i$ and D_i, must be instrumented in the 2SLS approach. As previously, SG_i and $deduct_i$ are instrumented by $income_i$ and edu_i through Probit regressions, leading to $\Pr(SG1_i) \equiv \Pr(SG_i = 1, deduct_i = 1)$ and $\Pr(SG2_i) \equiv \Pr(SG_i = 1, deduct_i = 0)$. Furthermore, edu_i and $income_i$ are also candidate instruments for D_i, because people with high income and high education level may have larger search costs, which may lead them to purchase insurance from a DOA. This is particularly the case when individuals purchase a new car, hence a third instrumental variable new_i, which indicates that the insured vehicle is less than three years old. Thus,

NT$$10,970$(US$365.67).

D_i is instrumented by:

$$\begin{aligned}
&\Pr(D_i = 1 | income_i, edu_i, new_i, X_i) \\
&= \Phi(\beta^3_{inc} income_i + \beta^3_{edu} edu_i + \beta^3_{new} new_i + \alpha^3 X_i).
\end{aligned} \tag{12}$$

Stage 2 of the 2SLS approach is now written as:

$$\begin{aligned}
&\Pr(SC_i = 1 | \Pr(SG1_i), \Pr(SG2_i), \Pr(D_i), RG_i, \Pr(D_i) * \Pr(SG1_i), \\
&\qquad \Pr(D_i) * \Pr(SG2_i), \Pr(D_i) * RG_i, X_i) \\
&= \Phi(\beta_{instr1} \Pr(SG1_i) + \beta_{instr2} \Pr(SG2_i) + \beta_D \Pr(D_i) + \beta_r RG_i \\
&\qquad + \beta_{Dinstr1} \Pr(D_i) * \Pr(SG1_i) + \beta_{Dinstr2} \Pr(D_i) * \Pr(SG2_i) + \beta_{Dr} \Pr(D_i) * RG_i + \beta X_i),
\end{aligned} \tag{13}$$

with $\Pr(SG1_i), \Pr(SG2_i)$, and $\Pr(D_i) \equiv \Pr(D_i = 1)$ being estimated at stage 1. In particular, we include interaction terms $\Pr(D_i) * \Pr(SG1_i)$ and $\Pr(D_i) * \Pr(SG2_i)$ in order to evaluate whether the conditional dependence between $SG1$ and SC and between $SG2$ and SC are comparatively higher in the DOA channel. The premium recouping effect and its interaction with the DOA channel are also taken into account through RG_i and $\Pr(D_i) * RG_i$, respectively.

At Stage 2 of the DGV approach, the explanatory variables include the dummy variables $SG1_i, SG2_i$ and the two estimated variables $\Pr(SG1_i), \Pr(SG2_i)$, with interaction terms to assess whether the conditional dependence between $SG1$ and SC and between

$SG2$ and SC is affected by the DOA channel. This is written as:

$$
\begin{aligned}
&\Pr(SC_i = 1|\Pr(SG1_i), \Pr(SG2_i), SG1_i, SG2_i, \Pr(D_i), RG_i, \Pr(D_i) * \Pr(SG1_i), \\
&\qquad\qquad \Pr(D_i) * \Pr(SG2_i), \Pr(D_i) * SG1_i, \Pr(D_i) * SG2_i, \Pr(D_i) * RG_i, X_i) \\
&= \Phi(\beta_{instr1}\Pr(SG1_i) + \beta_{instr2}\Pr(SG2_i) + \beta_{S1}SG1_i + \beta_{S2}SG2_i + \beta_D \Pr(D_i) + \beta_r RG_i \\
&\quad + \beta_{Dinstr1}\Pr(D_i) * \Pr(SG1_i) + \beta_{Dinstr2}\Pr(D_i) * \Pr(SG2_i) + \beta_{DS1}\Pr(D_i) * SG1_i \\
&\quad + \beta_{DS2}\Pr(D_i) * SG2_i + \beta_{Dr}\Pr(D_i) * RG_i + X_i\beta).
\end{aligned}
$$

The results are in Table 4.[27] The first column lists the estimated coefficients of the first stage regression for $\Pr(D)$: they confirm that individuals living in areas with high average income and high education level tend to purchase insurance through the DOA channel. This is also the case for the owners of vehicles that are less than three years old.

The 2SLS and DGV Probit regressions for SC are in the second and third columns. In the 2SLS Probit model, the estimated coefficients of $\Pr(SG1)$ and $\Pr(SG2)$ are 0.6522 and 1.7631, and they are significantly different from 0 at the 10% and 1% levels, respectively. The estimated coefficients of $\Pr(SG1*)\Pr(D\ )$and $\Pr(SG2)*\Pr(D)$ are 0.1067 and 0.3805, and they are significantly different from 0 at the 10% and 1 % level. All in all, this confirms that there is a significant conditional dependence between belonging to the suspicious group and filing claims during the suspicious period. This conditional dependence is even stronger among the insured who have purchased insurance through the DOA channel, and these effects are stronger for $SG2$ than for $SG1$. In other words, we may conclude that the fraud phenomenon associated with the claim date manipulation does exist, and that it is more severe among those individuals with deductible contracts and who have purchased insurance through the DOA channel, which confirms the prediction from Hypothesis 3.

[27] Here we have also checked the robustness of our IV method by using two sets of 2SLS-LPM and by checking that the null hypothesis of irrelevant model is rejected by the Durbin-Wu-Hausman test, the null hypothesis of exogenous instrumental variable cannot be rejected by the Anderson-Rubin test, and the null hypothesis of no-over identification cannot be rejected by the J test.

(Insert Table 4 here)

The third column of Table 4 corresponds to the DGV model. The results confirm our previous conclusions on the role of DOAs in the fraud process.[28] Furthermore, whatever the distribution channel, the $SG2$ coefficients are larger than their $SG1$ equivalents, which confirms that deductible contracts exacerbate fraudulent behaviors.

Calculation shows that the marginal effect of the estimated coefficients of $\Pr(D_i) * SG1_i$ and $\Pr(D_i) * \Pr(SG1_i)$ are equal to 0.1167 and 0.0439, which implies that, in the $SG1$ group, the probability of filing a claim during the suspicious period is 16.06% larger when policyholders have purchased insurance through the DOA channel than through another channel. Thus, if the expected cost of a fraudulent claim by an $SG1$ policyholder is US$199, as we have already estimated, then the expected fraudulent claim cost of such policyholders is $199 \times 0.1606 =$US$31.96 larger when insurance has been purchased through the DOA channel than through another channel. Similarly, the marginal effect of the estimated coefficients of $\Pr(D_i) * SG2_i$ and $\Pr(D_i) * \Pr(SG2_i)$ are equal to 0.1599 and 0.0545, thus with a 21.44% larger probability of filing a claim in the suspicious period for a member of the $SG2$ group who has purchased insurance through the DOA channel rather than through another channel. For an expected cost of fraudulent claims in the $SG2$ group equal to US$365.67, this amounts to an increase of US$78.40 in the expected cost of fraud when an $SG2$ policyholder takes out insurance from a DOA rather than through another distribution channel.

At the end, we may calculate the increase in fraud cost for each suspicious subgroup by comparison with the non-suspicious group, by weighting each subgroup with the corresponding number of policyholders. For example, $6,759$ policyholders in $SG1$ have taken out insurance through the DOA channel, with an expected increase in fraudulent claiming

[28] In particular, the estimated coefficients of interaction terms $\Pr(D) * \Pr(SG1)$ and $\Pr(D) * \Pr(SG2)$ are equal to 0.1997 and 0.2475, respectively, and they are significant at the 5% and 1% level. Similar conclusions are obtained for $\Pr(D) * SG1$ and $\Pr(D) * SG2$, with significance levels of 1% , respectively.

of US$108.71,[29] and hence a total additional cost of US$$734,771$. Similar calculations for the other cases yield the following results:

Increase in the cost of fraud	**DOA**	**Non-DOA**
$SG1$	$6,759 \times 108.71 = \$734,771$	$855 \times 76.75 = \$65,621$
$SG2$	$8,319 \times 253.63 = \$2,109,948$	$5,811 \times 175.23 = \$1,018,262$
$SG1 + SG2$	$\$2,844,719$	$\$1,083,883$

Overall, the suspicious policyholders in $SG1$ and $SG2$ who have purchased insurance through the DOA (respect. non-DOA) channel are at the origin of an increase in the cost of fraud that can be estimated at US$$2,844,719$ (respect.US$$1,083,883$), which corresponds to 13.72% (respect. 6.66%) of the premium written by this company for this line of business through DOAs (respect. non-DOAs).[30][31]

Remark 1: A legitimate question that may arise is whether the higher expected cost of claims in the DOA channel comes from fraudulent behaviors, as we have argued so far, or whether it rather reflects the fact that, on average, the individuals who take out insurance from DOAs have higher risks. An additional test, whose results are reported in the Appendix, shows that this not the case. In other words, the increase in claim costs

[29] According to the DGV model, the increase in the probability of filing a fraudulent claim by members of the $SG1$ subgroup who have purchased insurance through the DOA channel, in comparison with members of the non-suspicious group, is $0.1066 + 0.2791 + 0.0439 + 0.1167 = 0.5463$. Hence, the additional expected fraud cost: $199 \times 0.5463 =$US$108.71.

[30] In 2010, for this line of business, premiums written by this company through DOAs and non-DOAs amounted to USD20.73 and 16.27 million, respectively.

[31] Apart from these main results, Table 4 also provides two interesting by-products that are common to the 2SLS and DGV Probit models. Firstly, the estimated coefficient of RG_i is positive and significantly different from 0, at least at the 1% level, which confirms the existence of the premium recouping behavior. However, the estimated coefficients of the interaction term $RG_i * \Pr(D_i)$ have negative signs that are not significant. Thus, compared to other distribution channels, DOAs do not particularly help opportunistic policyholders to recoup premiums at the end of the policy year. Their behavior, as an act of collusion, rather focuses on the manipulation of the claim date. Secondly, the estimated coefficient of $female_i$ is positive and significant. This confirms that fraudulent behaviors may be widespread among those individuals who carefully manage their budget, since the declared gender of the owner of the car may be manipulated to take advantage of a lower premium. This is consistent with our previous observation made on Table 3.

is not an intrinsic characteristic of the distribution channel: it reflects the fraudulent behaviors of some policyholders (the suspicious groups) who may take advantage of the manipulation of claims, and this behavior is facilitated by DOAs.

6 Conclusion

This paper has focused attention on the policyholder-service provider coalition in insurance mechanisms: how it can affect the credibility of claim auditing, how several patterns of fraud may emerge in the car insurance market, and how service providers and policyholders may draw benefit from such a coalition. The important role of car dealers in Taiwan provides an exceptional opportunity to analyze this interaction between insurer, policyholder and provider.

Indeed, the economic analysis of insurance fraud is usually based on a very abstract picture of claims fraud (filing a fraudulent claim although no accident has occurred, or exagerating a claim), but in practice understanding insurance fraud often requires a much more specific analysis of the claims fraud process. The Taiwan case offers such a possibility, with fraud frequently taking place through the manipulation of the claim's date in order to avoid a penalty from the bonus-malus system and to reduce the burden of a second deductible, should another accident occur.

We hope to have brought convincing evidence that the intertemporal manipulation of claims is actually a significant determinant of insurance fraud in Taiwan. In particular, policyholders with deductible contracts who intend to renew their policies (the suspicious group) have a larger propensity to defraud than other policyholders, by postponing their claims until the last month of the policy year, and possibly by merging two events into a single claim. Consequently, there is an increase in the average cost of first claims filed by the suspicious group in the last month of the policy year. Furthermore, the collusion between policyholders and DOAs is a crucial mechanism that contributes to the development of fraud in the Taiwanese car insurance market.

The size of claim manipulation in the Taiwan insurance market and the role of DOAs are so significant that it is hard to believe that insurers are unaware of them. Informal exchanges with the industry confirm that this is the case. Of course, there may be different views of the underlying mechanisms, varying from granting small advantages to policyholders in order to build customers loyalty and closing ones eyes to false small claims - i.e., the end of year "car wash" and the recouping money behavior, respectively - up to organized large scale insurance fraud through claim manipulation and collusion with car repairers. All these components of customers misbehaviors are likely to coexist. The lack of response of Taiwan insurers concerned is more striking. Our analysis suggest that it is in fact very difficult to incentivize DOAs through contractual mechanisms so that fighting fraud would be in their own interest. The main revenue of DOAs comes from the sales of new vehicles and from the activity of their repair shops. Although DOAs are major intermediaries in the Taiwan insurance market, selling insurance is for most of them mainly a way to improve their relationships with car buyers, it is not their core business. Reducing the intensity of fraud through the DOA channel would in fact require a structural reorganization that would be very costly to insurers. Integrating DOAs as part and parcel of the firm (i.e., regrouping the insurance activity and the sales of cars within a common holding company) might completely change the story. It would reduce the bargaining power of DOAs and, presumably, it would also allow insurers to reduce the asymmetry of information with their DOAs by implementing regular cost reviews, instead of triggering costly audits on a case by case basis, and with the insurer in a position of weakness. Other Taiwanese insurers have made that choice. The size of fraud in the Taiwan automobile insurance market should persuade insurers that comparing the costs and benefits of these decentralized and integrated schemes is of utmost importance.

References

Chiappori, P.A. and Salanié, B. (2000), "Testing for asymmetric information in insurance markets", *Journal of Political Economy*, 108(1): 56–78.

Dionne, G., and Gagné, R. (2001), "Deductible contracts against fraudulent claims: evidence from automobile insurance", *Review of Economics and Statistics*, 83, 290–301.

Dionne, G. and R. Gagné R. (2002), "Replacement cost endorsement and opportunistic fraud in automobile insurance", *Journal of Risk and Uncertainty*, 24: 213-230.

Dionne, G., F. Giuliano and P. Picard (2009a), "Optimal auditing with scoring: theory and application to insurance fraud", *Management Science*, 55, 58-70.

Dionne, G., C. Gouriéroux and C. Vanasse (1997), "The informational content of household decisions with application to insurance under adverse selection", *Manuscript*. Montréal: Ecole Hautes Etudes Commerciales.

Dionne, G., C. Gouriéroux and C. Vanasse (2001), "Testing for evidence of adverse selection in the automobile insurance market : a comment", *Journal of Political Economy*, 109(2) : 444-453

Dionne G., P. St-Amour and D. Vencatachellum (2009b), "Asymmetric information and adverse selection in Mauritian slave auctions", *Review of Economic Studies*, 76 (4): 1269-1295.

Dionne, G., M. La Haye and A. S. Bergeres (2015) "Does asymmetric information affect the premium in mergers and acquisitions?", *Canadian Journal of Economics*, 48(3), 819-852.

Dionne, G. and Wang, K. C. (2013), "Does opportunistic fraud in automobile theft insurance fluctuate with the business cycle?", *Journal of Risk and Uncertainty*, 47, 67-92.

Fukukawa, K., C. Ennew and S. Diacon (2007), "An eye for an eye : investigating the impact of consumer perception of corporate unfairness on aberrant consumer behavior", in *Insurance Ethics for a more Ethical World (Research in Etical Issues in Organizations)*, edited by P. Flanagan, P. Primeaux and W. Ferguson, Vol.7, 187-221, Emerald Group Publishing Ltd.

Gouriéroux, C., A. Monfort (1995), "Statistics and econometric models", *Cambridge University Press*, 2: 458–475.

Harris, L.C. and K. Daunt (2013), "Managing customer misbehavior: Challenges and strategies", *Journal of Services Marketing*, 27(4): 281-293.

Li, C-S., C-C. Liu and S-C. Peng (2013), "Expiration dates in automobile insurance contracts : The curious case of last policy month claims in Taiwan", *Geneva Risk and Insurance Review*, 38(1): 23-47.

Mayer, D. and Jr. C. S. Smith (1981), "Contractual provisions, organizational structure, and conflict control in insurance markets", *Journal of Business*, 54: 407-434.

Miyazaki, A.D. (2009), "Perceived ethicality of insurance claim fraud : do higher deductibles lead to lower ethical standards", *Journal of Business Ethics*, 87(4):589-598.

Murthy, D.N.P. and I. Djamaludin (2002), "New product warranty: a literature review", *International Journal of Production Economics*, 79, 231-260.

Pao, T.I., L.Y. Tzeng and K.C. Wang (2014), "Typhoons and opportunistic fraud: claim patterns of automobile theft insurance in Taiwan", *Journal of Risk and Insurance*, 81(1): 91-112.

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

Tennyson, S. (1997), "Economic institutions and individual ethics : a study of consumer attitudes toward insurance fraud", *Journal of Economics Behavior and Organization*, 32:247-265.

Tennyson, S. (2002), "Insurance experience and consumers'attitudes toward insurance fraud", *Journal of Insurance Regulation*, 21(2):35-55.

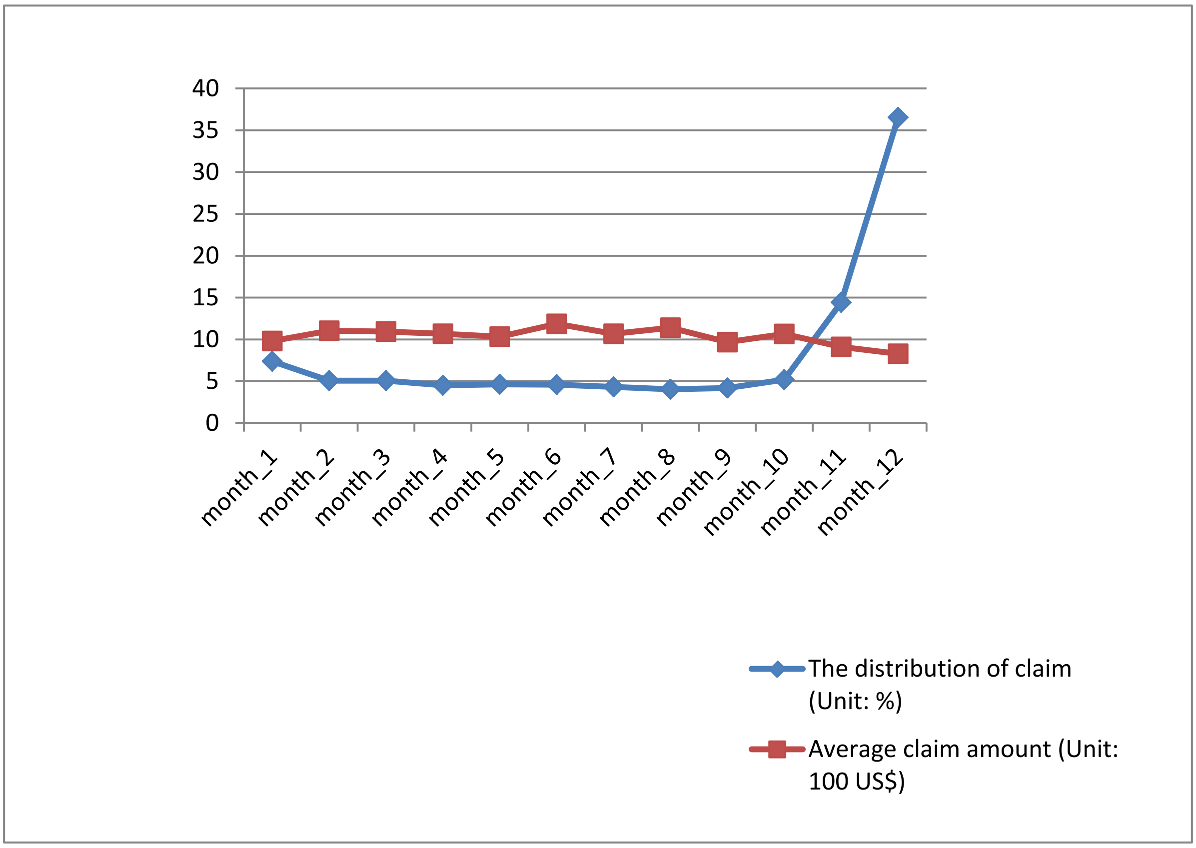

Figure 1: Distribution of claims and average claim cost (first claims) in the policy year

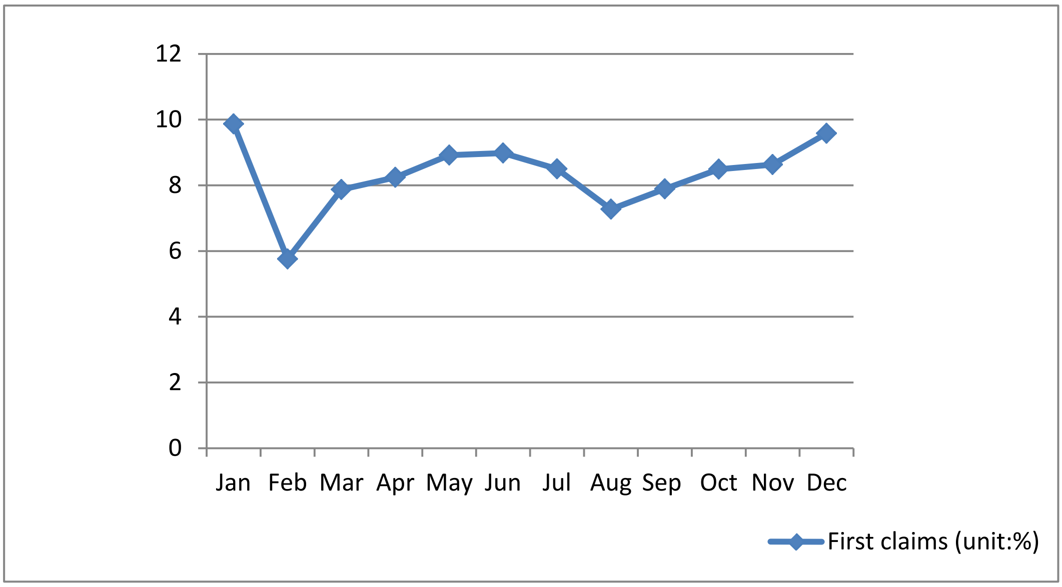

Figure 2: Distribution of the first claims among calendar months

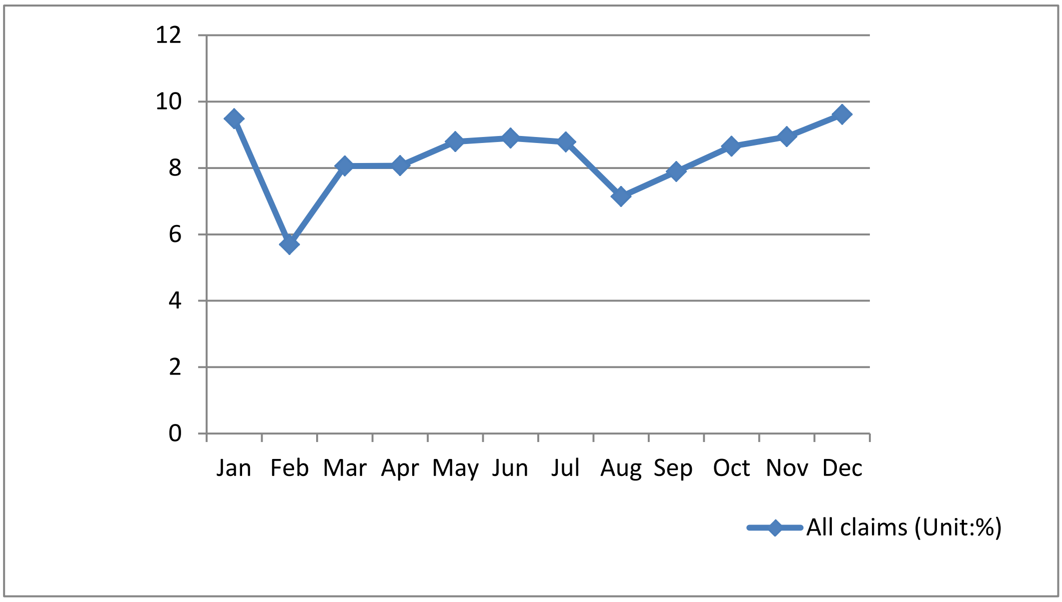

Figure 3: Distribution of all claims among twelve calendar months

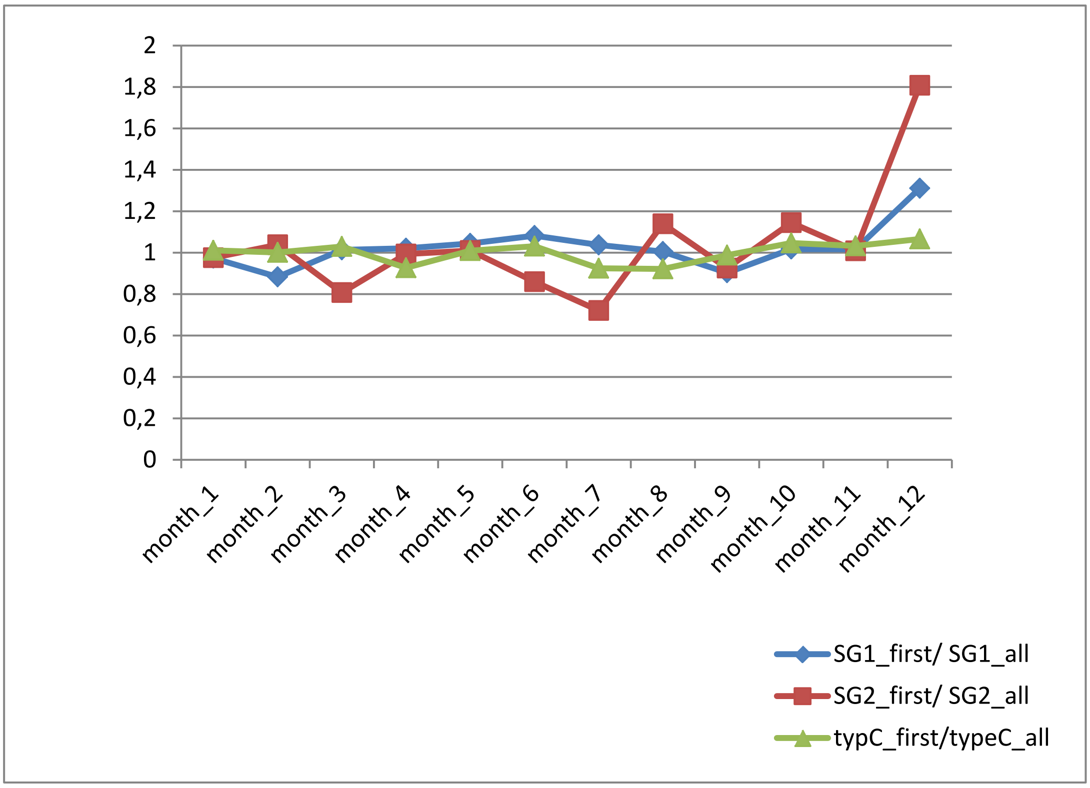

Figure 4: Average cost of first claims / Average cost of all claims
Comparing the suspicious group and type C contracts

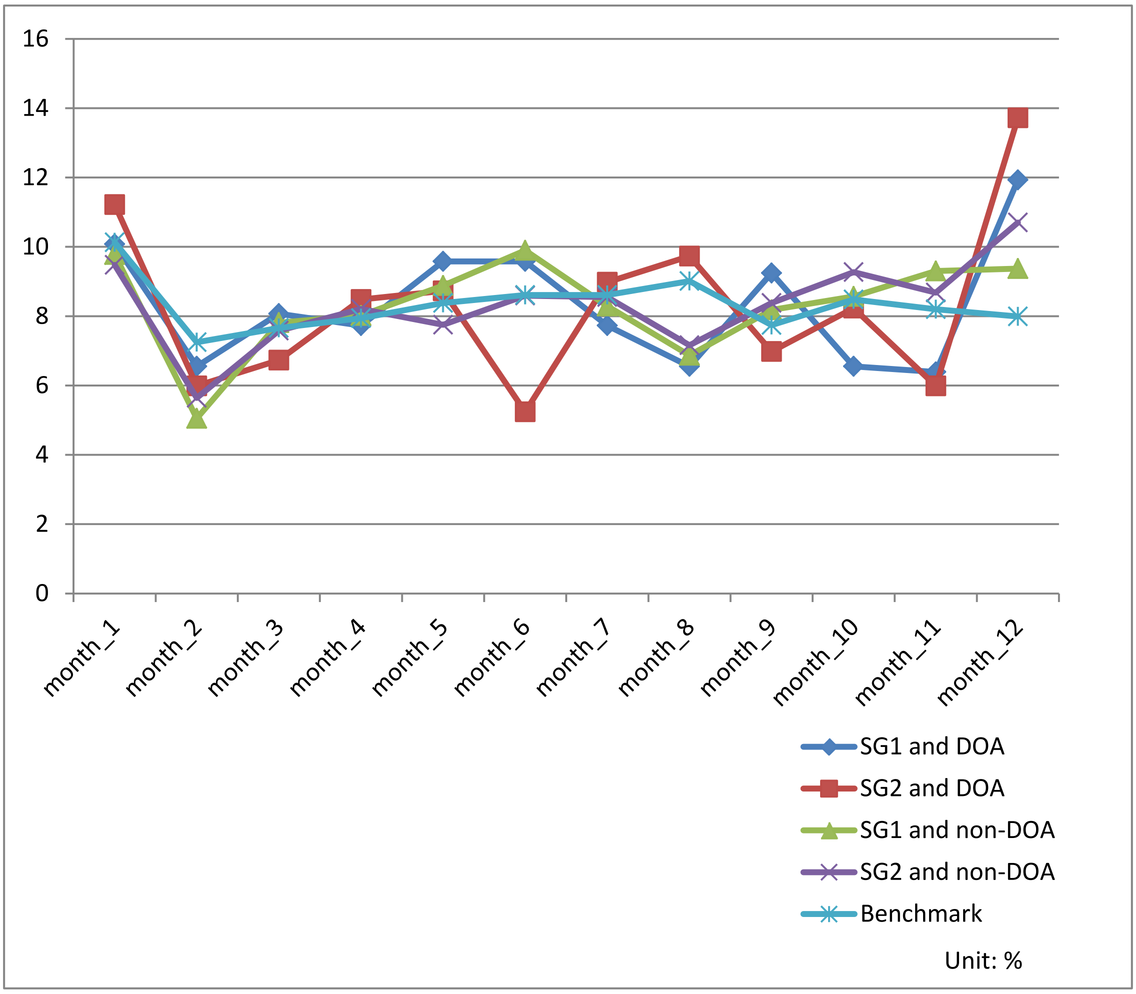

Figure 5: Distribution of claims during the policy year

Table 1: Variable definitions

Variable	Definition
Explained variable:	
SG	Dummy variable equal to 1 when the insured belongs to the "suspicious group",[1] and 0 otherwise.
SG1	Dummy variable equal to 1 when the insured belongs to "suspicious group 1",[2] and 0 otherwise.
SG2	Dummy variable equal to 1 when the insured belongs to "suspicious group 2",[3] and 0 otherwise.
deduct	Dummy variable equal to 1 when the insured has taken out a deductible contract, and 0 otherwise.
SC	Dummy variable equal to 1 when the insured has filed his or her first claim during the suspicious period (in the last policy month), 0 otherwise.
First group of explanatory variables: underwriting and pricing factors	
female	Dummy variable equal to 1 if the insured is a female, 0 otherwise.
age2025	Dummy variable equal to 1 if the insured is in the 20-25 age group, 0 otherwise.
age2530	Dummy variable equal to 1 if the insured is in the 25-30 age group, 0 otherwise .
age3060	Dummy variable equal to 1 if the insured is in the 30-60 age group, 0 otherwise.
ageabv60	Dummy variable equal to 1 if the insured is older than 60, 0 otherwise.
carage0	Dummy variable equal to 1 when the car is less than one year old, 0 otherwise.
carage1	Dummy variable equal to 1 when the car is two years old, 0 otherwise.
carage2	Dummy variable equal to 1 when the car is three years old, 0 otherwise.
carage3	Dummy variable equal to 1 when the car is four years old, 0 otherwise.
carage4	Dummy variable equal to 1 when the car is more than four years old, 0 otherwise.
veh_m	Dummy variable equal to 1 when the capacity of the insured car is between 1800 and 2000 c.c., 0 otherwise.
veh_l	Dummy variable equal to 1 when the capacity of the insured car is larger than 2000, 0 otherwise.
tramak_j	Dummy variable equal to 1 when the brand of the insured car is *j*, with *j*=*n*, *f*, *h*, *t*, *c*,

[1] The "suspicious group" (*SG*) includes the individuals who purchased type A or B contract and renewed their contract with the same insurance company. The counter group for SG includes the policyholders who purchased type C contract or who did not renew their contract with the same insurance company.

[2] The "suspicious group 1" (*SG*1) includes the *SG* group policyholders with no-deductible contract. The counter group for *SG1* includes the policyholders who purchased type C contract and who did not renew their contract with the same insurance company, or who belong to *SG2*.

[3] The "suspicious group 2" (*SG*2) includes the policyholders with deductible contract who renewed their contract with the same insurance company. The counter group for *SG2* includes the policyholders who purchased type C contract and who did not renew their contract with the same insurance company, or who belong to *SG1*.

	and 0 otherwise.[4]
sedan	Dummy variable equal to 1 when the car is a sedan and is for non-commercial or for long-term rental purposes, and 0 otherwise.[5]
logprem	Logarithm of the premium (in US dollars) of the contract in the current contract year.
bonus	Bonus-malus coefficient used to calculate the premium in the current contract year. It is a multiplier on the premium. Hence, it is a discount if it is smaller than 1 and it is a penalty if it is larger than 1.
Explanatory variables (second group):	
income	Average income in the policyholder's residential area.
edu	Percentage of inhabitants with a PhD or a master degree in the policyholder's residential area.
new	Dummy variable equal to 1 when the car is less or equal to three year old, and 0 otherwise.
D	Dummy variable equal to 1 if the insurance contract is sold through the DOA channel, and 0 otherwise.
A	Dummy variable equal to 1 for a type A contract, and 0 otherwise.
B	Dummy variable equal to 1 for a type B contract, and 0 otherwise.[6]
RG	Dummy variable equal to 1 when the insured belongs to the "recoup group",[7] and 0 otherwise.

[4] The counter group for *tramak_j, j= n, f, h, t, c* corresponds to brands other than Nissan, Ford, Honda, Toyota, and China.

[5] The counter group includes cars that are not small sedans, for example small or large trucks, cargos…etc.

[6] The counter groups for *A* and *B* are type C contracts.

[7] The "recoup group" includes the policyholders who are covered by type A or B contracts and who did not renew their contract or renewed it for only one year.

Table 2-1: Structures of the whole sample and of the sub-sample with claims

	Whole sample (A)	Sub-sample with claims (B)	**Difference (B)-(A)**
claim	0.4586		
SC		0.1751	
RG	0.1980	0.2847	0.0897***
deduct	0.3602	0.5670	-0.2068***
A	0.0103	0.0152	0.0049***
B	0.3882	0.6742	0.2860***
SG	0.2625	0.4332	0.1707***
SG1	0.1177	0.1517	0.0340***
SG2	0.1448	0.2815	0.1367***
D	0.5078	0.6216	0.1138***
female	0.7118	0.7392	0.0274***
age2025	0.0030	0.0028	-0.0002
age2530	0.0342	0.0370	0.0028***
age3060	0.8947	0.8951	0.0004
ageabv60	0.0679	0.0651	-0.0028**
carage0	0.2192	0.2756	0.0566***
carage1	0.1381	0.1891	0.0510***
carage2	0.0915	0.1025	0.0110***
carage3	0.1109	0.1062	-0.0047***
carage4	0.0986	0.0842	-0.0144***
veh_m	0.2875	0.2626	-0.0249***
veh_l	0.2692	0.2849	0.0157***
sedan	0.9166	0.9325	0.0159***
logprem	5.8240	6.2096	0.3856***
bonus	0.7180	0.8028	0.0848***
No of obs.	109,461	50,194	

Notes

(1) The information on car brands in the two samples is not reported for confidentiality reasons.

(2) ***,** and * indicate 1%, 5% and 10% significance levels, respectively.

Table 2-2: Structure of the DOA and non-DOA subsamples

	DOA (A)	Non-DOA (B)	Difference (A)-(B)
SC	0.2299	0.0849	0.1450***
RG	0.1373	0.3744	-0.2371***
deduct	0.5461	0.6012	-0.0551***
A	0.0145	0.0164	-0.0019*
B	0.7190	0.6007	0.3183***
SG	0.4832	0.3509	0.1323***
SG1	0.2166	0.0450	0.1716***
SG2	0.2666	0.3059	-0.0393***
female	0.7597	0.7052	0.0546***
age2025	0.0027	0.0036	-0.0009
age2530	0.0393	0.0333	0.0060***
age3060	0.8974	0.8913	0.0062**
ageabv60	0.0603	0.0730	-0.01267***
carage0	0.3916	0.0850	0.3066***
carage1	0.2203	0.1379	0.0824***
carage2	0.1021	0.1033	-0.0012
carage3	0.0890	0.1344	-0.0454***
carage4	0.0601	0.1239	-0.0638***
veh_m	0.2287	0.3183	-0.0896***
veh_l	0.2878	0.2800	0.0078*
tramak_n	0.0058	0.0098	-0.0039***
tramak_f	0.0377	0.0763	0.0386***
tramak_h	0.0561	0.0994	-0.0433***
tramak_t	0.6464	0.3388	0.3076***
tramak_c	0.0070	0.0601	-0.0532***
sedan	0.9492	0.9050	0.0442***
logprem	6.4674	5.7862	0.6812
bonus	0.8708	0.6909	0.1799***
No of obs.	31,203	18,991	

Note

(1) ***,** and * indicate 1%, 5% and 10% significance levels, respectively.

Table 3: Empirical evidence of fraud

	First stage (bivariate Probit)		Second stage	
	SG	***deduct***	**2SLS-*Probit***	**DGV-*Probit***
constant	-1.6092***	-2.6576***	-4.6113***	-4.2856***
Pr(*SG*1)			0.6110*	0.6094*
Pr(*SG*2)			0..8021***	0.7809***
***SG*1**				1.2211*
***SG*2**				1.7110***
income	-8.93E-06***	2.71E-05***		
edu	0.2733*	-1.7944***		
RG	-0.4377***	0.4810***	0.3937***	0.4524***
female	0.1525 ***	-0.0913***	0.0695***	0.0559***
age2530	0.4104	0.3272	0.4801***	0.4548***
age3060	0.5899	0.2349	0.5643***	0.5269***
ageabv60	0.6027	0.2977	0.4463***	0.4199***
carage0	0.7029***	0.1488***	0.8574***	0.8735*
carage1	0.4539***	0.0757**	0.0450	0.0521
carage2	0.3554*	0.0981***	0.0097	0.0152
carage3	0.2864*	0.0971***	-0.0087	-0.0059
carage4	0.1370	0.0376	-0.0149	-0.0134
veh_m	0.0228	0.0338	-0.1252***	-0.1100***
veh_l	0.3493***	0.3530***	-0.3165***	-0.2838***
sedan	0.2085***	0.2249***	-0.1735***	-0.1568***
logprem	0.6398***	0.0623***	0.5673***	0.4937***
bonus	-0.8046***	-0.0520***	-1.0049***	-0.9060***
Pseudo R^2	0.5523		0.2035	0.2043

Notes

(1) Pr($SG1_i$) and Pr($SG2_i$) are the estimated probabilities of belonging to the suspicious groups *SG*1 and *SG*2, respectively, calculated at the first stage, that is Pr($SG1_i$) = $Prob(SG_i$=1, $deduct_i$=0), and Pr($SG2_i$) = $Prob(SG_i$=1, $deduct_i$=1). In the DGV-probit model, *SG*1 and *SG*2 are dummy variables for belonging to the suspicious groups *SG*1 and *SG*2, respectively.

(2) In all the above regressions, we have also controlled for the brand of the insured car. This is not reported for confidentiality reasons.

(3) ***,** and * indicate 1%, 5% and 10% significance levels, respectively.

(4) We have also performed two sets of the 2SLS-LPM to confirm the validity of our IV model. In both sets, the null hypothesis of irrelevant model is rejected by the Durbin-Wu-Hausman test, the null hypothesis of exogenous instrumental variable cannot be rejected by the Anderson-Rubin test, the null hypothesis of no over identification cannot be rejected by the *J* test.

Table 4: **Empirical evidence of fraud through DOAs**

	First stage	**Second stage**	
	D	**2SLS-*Probit***	**DGV-*Probit***
constant	-2.8831***	-4.8600***	-4.4864***
Pr(*SG1*)		0.6522*	0.4844*
Pr(*SG2*)		1.7631***	0.6105***
***SG*1**			1.2686***
***SG*2**			1.5675***
Pr(*D*)		0.1879	-0.1702
Pr(*D*)*Pr(*SG1*)		0.1067*	0.1997*
Pr(*D*)*Pr(*SG2*)		0.3805***	0.2475**
Pr(*D*)*SG1*			0.5306***
Pr(*D*)*SG2*			0.7269***
income	1.72E-05***		
edu	3.6226***		
new	0.2170***		
RG	0.2383***	0.4328***	0.4814***
Pr(*D*)*RG*		-0.0183	-0.0108
female	0.0934***	0.0787***	0.0605***
age2530	0.1786	0.4987***	0.4659***
age3060	0.4100***	0.6037***	0.5597***
ageabv60	0.3697***	0.4825***	0.4514***
carage0	0.9477***	0.9566***	1.0231***
carage1	0.5248***	0.1089**	0.1538***
carage2	0.3694***	0.0536	0.0860*
carage3	0.2418*	0.0144	0.0350
carage4	0.1416	-0.0023	0.0093
veh_m	-0.1285***	-0.1322***	-0.1138***
veh_l	-0.2397***	-0.3403***	-0.2962***
sedan	0.0069	-0.1788***	-0.1502***
logprem	0.3610***	0.6011***	0.5016***
bonus	0.2463 ***	-0.9725***	-0.8045***
Pseudo R^2	0.2390	0.2040	0.2054

Same notes as in Table 3

Appendix

Available online - Not to be published

1 A model of insurance fraud with policyholder-car repairer collusion

For the sake of brevity, Section 3 of the paper is limited to a simple insurance fraud model, in which the choice of the insurance contract and of the distribution channel are left unexplained. It is just mentioned that the level of the deductible reflects the policyholder's risk aversion, and that individuals have some preferences for a specific distribution channel. Furthermore, the collusion process involving policyholders and car repairers is not precisely described in this model. We here develop an integrated model where these issues are explicitely analyzed. The model presented in the paper may be viewed as a simplified version of this integrated model.

1.1 Notations

We consider an economy with a competitive insurance market, in which automobile insurance can be purchased either through car dealers who act as insurance agents (DOAs) or through independent insurance agents. Car dealers also own auto repair shops. Accidents may be minor or severe, with repair costs ℓ and 2ℓ whatever the car repairer, for minor and serious accidents respectively, and also an uninsurable loss ε per accident.[1] Insurance policies consist of a premium P and possibly a deductible d for each accident.[2] Insurance pricing includes

[1] Assuming that the insurable costs of severe accidents exactly double those of minor accidents simplifies the notations of the model. We could more generally assume that severe accidents cost more than minor accidents. The repair shop market is competitive, so that policyholders can let their car be repaired at competitive price ℓ or 2ℓ whatever the insurance distribution channel. The uninsurable loss ε corresponds to earnings losses, time value, daily life disruption or stress incurred in the case of an accident. This loss does not play a significant role in our theoretical analysis, but it makes it possible for some individuals to choose a deductible contract while others prefer a full coverage contract (in what follows, the type 1 and 2 individuals respectively), which will fit our empirical analysis of the Taiwan automobile insurance market.

[2] The fact that deductibles are per accident follows the usual practice of car insurance companies (of course not only in Taiwan), although it does not correspond to an optimal insurance contract design. This feature of automobile insurance probably reflects the increase in transaction costs that would be induced by aggregate deductibles over the whole period covered by the contract. For notational simplicity, we assume that the deductible is the same for the first and second claims. In Taiwan, second claims have larger deductibles than the first one that occurred during the same policy year, which may be viewed as an incentive device in a moral hazard setting (see Li et al, 2007).

constant proportional loading σ, and insurers may offer different policies through car dealers and through other distribution channels.

Each individual may suffer from 0, 1 or, at most, 2 accidents during the policy year. Let π_1 and π_2 be respectively the probability of 1 and 2 accidents, with $0 < \pi_1 + \pi_2 < 1$. Each accident is minor with probability q_m and severe with probability q_s, with $q_m + q_s = 1$. The policy year is divided in two sub-periods, which are called the non-suspicious period (NSP) and the suspicious period (SP), respectively, because, as we will see, filing a claim in SP may be a signal of fraud. Any accident occurs in SP with probability μ, with $0 < \mu < 1$.[3]

There are two types of individuals with the same initial wealth w: type 1 has a larger degree of absolute risk aversion than type 1. Let w_f be the individual's final wealth. $u_h(w_f)$ denotes the type h von Neumann-Morgenstern utility function (with $h = 1$ or 2), and we assume $u_h' > 0$ and $u_h'' < 0$, and

$$-\frac{u_1''(w_f)}{u_1'(w_f)} > -\frac{u_2''(w_f)}{u_2'(w_f)},$$

for all w_f. Let λ_h be the proportion of type h individuals, with $\lambda_1 + \lambda_2 = 1$. Car repairers are risk neutral.

We also assume that individuals have differentiated preferences between purchasing insurance through a car dealer or through an independent agent. In particular, individuals who have high search costs may prefer to purchase insurance through car dealers because often purchasing a new car goes together with taking out a new insurance policy. This is modelled as in a Hotelling game. Both types of individuals are uniformly located on the interval $[0, 1]$. A representative DOA and another representative independent insurance agent are located at the extremities of the $[0, 1]$ segment: the DOA is at $x = x_D = 0$ and owns a repair shop, while the the other distribution channel is at $x = x_A = 1$.

Purchasing insurance entails a search disutility which is proportional at rate t to the distance covered to 0 and 1 according to the distribution channel. Thus, the expected utility of a type h customer located at $x \in [0, 1]$ with contract (P, d) is

$$\overline{u}_h(P, d) - t\,|x - x_i|\,,$$

[3]For instance, if SP corresponds to the last policy month, and if accidents are uniformly distributed over the policy year, then $\mu = 1/12$.

where

$$\overline{u}_h(P,d) \equiv (1-\pi_1-\pi_2)u_h(w-P)+\pi_1 u_h(w-P-d-\varepsilon)+\pi_2 u_h(w-P-2d-2\varepsilon), \quad \text{(A-1)}$$

for $h=1$ or 2, with $i=D$ if that customer purchases insurance through the representative DOA and $i=A$ if he goes through the other distribution channel.

Type 1 individuals have a larger propensity to purchase insurance coverage than type 2 since they are more risk averse. Because of these differentiated preferences, insurers offer menus of contracts. Let (P_{ih}, d_{ih}) be the insurance contract that is taken out by type h individuals, with $i=A$ or D according to the distribution channel.

1.2 The fraud mechanism

Fraud is analyzed as the behavior of oppportunistic policyholders who delay their claims to SP, with the complicity of a car repairer. We consider a very simple form of the opportunistic policyholder-car repairer collusive game. The policyholder makes a take-it-or-leave-it offer to the car repairer in which he offers to pay a fixed amount G to the repairer and he keeps the residual part of the collusive gain. Because of the bonus-malus system, type 1 and 2 policyholders may commit such a fraud in order to avoid paying a higher premium during the next policy year: v denotes the discounted value of the savings in future insurance premiums induced by such a bonus-malus fraud. Only those individuals who plan to renew their contract with the same insurer may profit from such a bonus-malus fraud. We assume that they make up a proportion $\delta \in (0,1)$ of the policyholders (whatever their type).

We also assume that postponing a minor claim requires that another minor loss actually occurs during the same policy year, so that the total losses may be presented as the outcome of a single severe accident. Policyholders also get an additional advantage from fraud by reducing the retained cost from $2d_{ih}$ to d_{ih}. Thus, if fraud has been committed and is not detected, the collusive gain is $d_{ih}+v$ or d_{ih} if the claim is filed in SP or NSP, respectively. It is shared between repairer and policyholder as amounts G_{SP} and $d_{ih}+v-G_{SP}$ or G_{NSP} and $d_{ih}-G_{NSP}$, where G_{SP} and G_{NSP} denote the transfer to the repairer when the fraudulent claim is filed in SP and NSP, respectively .

Thus, if a minor accident occurs in NSP, then the policyholder may decide not to immediately file a claim for this accident. Two possible cases are then possible.[4] If another minor accident occurs later during the same policy year, then the policyholder may file a single large claim for the two accidents (called a "fraudulent claim" in what follows), which requires collusion with a car repairer. Auditing large claims allows the insurer to detect such instances of fraud. We denote as c_i the cost of an audit when insurance is purchased from $i \in \{D, A\}$. The fact that the car dealer owns the repair shop makes collusion all the easier. Thus, we assume that auditing claims is more costly (or, put differently, it is more difficult to establish colluders' fraud) when insurance is purchased from D than from A.[5] We thus assume $c_D > c_A$. If there is no other minor accident, then the insurer considers that any late claim (for the first accident) is invalid and is dismissed. If a policyholder is caught filing a fraudulent claim through a collusive agreement with the repairer, then he has to pay a fine B, and he does not receive a indemnity, and the repairer pays a fine B'. For simplicity, fines are determined exogenously by law. They are entirely paid to the State budget and are not part of the insurer's income. They may also be interpreted as the litigation costs incurred by the policyholder and the car repairer when fraud is discovered.[6]

1.3 Fraud-audit interaction

Let $\alpha_{ih} \in [0, 1]$ be the fraud rate of type $h \in \{1, 2\}$ individuals who purchase inurance from $i \in \{A, D\}$. This is the fraction of type h policyholders who decide not to immediately file a claim when a minor accident occurs in NSP, hoping for a future collusive agreement with a car repairer, should another minor accident occur in SP.[7] Let $\widehat{\pi} = \pi_2/(\pi_1+\pi_2)$ be the probability

[4] Bear in mind that in what follows we neglect the possibility of more than two accidents for the same policyholder. We also assume that there are only two types of acccidents (minor or severe) with repair costs of ℓ and 2ℓ, respectively. Thus, we do not contemplate the possibility of presenting, say, a minor accident and a serious accident as an extreme accident with cost 3ℓ. In other words, the falsification of claims only consists of announcing one single severe accident instead of two minor accidents.

[5] For example, in the DOA case, the hidden transfer G may take the form of a promise to purchase a new car in the near future.

[6] In practice, when fraud is discovered, the policyholder-repairer coalition has some bargaining power that may allow its members to escape the penalties. This is particularly the case when insurance has been purchased from a DOA, because the latter is in a position to threaten the insurer with redirecting its (presumably large) customer base toward another insurer. This is another reason why deterring fraud may be more difficult when insurance has been taken out through a DOA than through a standard agent. The effects of agents' bargaining power on the enforcement of fraud penalties is analyzed in Section 1.6 of this Appendix, and for the sake of presentation simplicity is not taken into account here.

[7] We may check that policyholders would not take advantage of colluding with a repairer if a second accident occurs during the non-suspicious period.

of having a second accident, conditionally on the occurence of a first accident in NSP.[8] Such an accident will occur in SP with probability μ, and it will be minor with probability q_m. Thus, if a first minor acccident occurs in NSP, then a future collusive agreement with a car repairer will be possible with probability $q_m\mu\widehat{\pi}$. The audit of serious claims may detect such fraud. These audits are triggered with probability $\beta_{ih} \in [0,1]$.[9] In short, α_{ih} and β_{ih} for $i = A, D$ and $h = 1, 2$ are the policyholder's and insurer's strategies, respectively.

The expected utility of a type h policyholder who does not immediately file a claim after a first (minor) accident in NSP is written as:[10]

$$
\begin{aligned}
Eu_{ih}^F \quad = \quad & q_m\widehat{\pi}\mu[(1-\beta_{ih})u_h(w - P_{ih} - d_{ih} - 2\varepsilon + v - G_{SP}) \\
& +\beta_{ih}u_h(w - P_{ih} - 2\ell - 2\varepsilon - G_{SP} - B)] \\
& q_m\widehat{\pi}(1-\mu)[(1-\beta_{ih})u_h(w - P_{ih} - d_{ih} - 2\varepsilon - G_{NSP}) \\
& +\beta_{ih}u_h(w - P_{ih} - 2\ell - 2\varepsilon - G_{NSP} - B)] \\
& +\widehat{\pi}q_s u_h(w - P_{ih} - \ell - d_{ih} - 2\varepsilon) + (1-\widehat{\pi})u_h(w - P_{ih} - \ell - \varepsilon).
\end{aligned}
$$

This formula may be interpreted as follows. If the policyholder does not immediately file a claim after a minor accident in NSP, he will have the opportunity to defraud if there is another minor accident. This second accident will occur in SP or NSP with probability $q_m\widehat{\pi}\mu$ or $q_m\widehat{\pi}(1-\mu)$, respectively. In these cases, either the claim is audited or not, respectively with probabilities β_{ih} and $1-\beta_{ih}$. If there is no audit, then the policyholder receives either $d_{ih} + v - G_{SP}$ or $d_{ih} - G_{NSP}$, which is his share of the collusive deal, in addition to his status quo net wealth $w - P_{ih} - 2d_{ih}$ (i.e., the policyholder's wealth in the case of two accidents without fraud). If there is an audit, then no indemnity is paid by the insurer, and the policyholder pays the fine B and does not recoup his side-payment G_{SP} or G_{NSP}. If no

[8] For simplicity, we do not condition this probability on the exact date at which the first accident occurs. In other words, we consider the non-suspicious period as a whole.

[9] We will assume that all serious claims (for i and h given) are audited with the same probability β_{ih}. In other words, the audit frequency is not conditional on whether the claim is filed during the suspicious or non-suspicious period. This seems to be a realistic assumption insofar as the beginning of the policy year varies across individuals, and conditioning auditing on the date of the claim in the policy year of each individual would probably entail substantial transaction costs. Be that as it may, concentrating audits on the suspicious period individual by individual would increase the efficiency of the fraud deterrence mechanism, but this would not qualitatively affect our conclusions.

[10] The formula would be almost unchanged if the first accident also occurs in the suspicious period. In such a case, the gain from collusion would be lower (v should be replaced by a lower collusive gain v') because the advantage from bonus-malus fraud would be lower. Consequently, defrauding by filing a single claim for two minor accidents in the suspicious period does not occur for the equilibrium audit strategy.

fraudulent claim can be filed, then the late claim is dismissed: another accident occurs and it is severe with probability $\widehat{\pi}q_s$, and there is no other accident with probability $1-\widehat{\pi}$. In both cases, no insurance indemnity is paid for the first claim.

If the policyholder immediately files a claim after his first minor accident, then his expected utility (after this first accident) is

$$Eu_{ih}^N = \widehat{\pi}u_h(w - P_{ih} - 2d_{ih} - 2\varepsilon) + (1-\widehat{\pi})u_h(w - P_{ih} - d_{ih} - \varepsilon).$$

The policyholder is willling to defraud by making side-payment G_{SP} or G_{NSP} to the car repairer if $Eu_{ih}^F \geq Eu_{ih}^N$, that is if $\beta_{ih} \leq \Psi_h(P_{ih}, d_{ih}, G_{SP}, G_{NSP})$, where function $\Psi_h(.)$ is such that[11] $\Psi_h(P_{ih}, d_{ih}, G_{SP}, G_{NSP}) < 0$ if $d_{ih} = v = 0$, which reflects the obvious fact that no audit is required to dissuade fraud if the defrauders have nothing to earn by postponing their claims. If d_{ih} and/or v are large enough for auditing to be necessary, then we have $\Psi_h(P_{ih}, d_{ih}, G_{SP}, G_{NSP}) \in (0,1)$ and $\partial\Psi_h/\partial G_{SP}, \partial\Psi_h/\partial G_{NSP} < 0$. We focus on this case in what follows. The repairer agrees to collude in SP and NSP if his expected gain from collusion is positive, that is, if $G_{SP} - \beta_{ih}B' \geq 0$ and $G_{NSP} - \beta_{ih}B' \geq 0$.The optimal side-payment offer from the policyholder to the car repairer is thus $G_{SP} = G_{NSP} = \beta_{ih}B'$. The policyholder is indifferent between defrauding (through an optimal hidden agreement with the car repairer) and not defrauding if $\beta_{ih} = \Psi_h(P_{ih}, d_{ih}, \beta_{ih}B', \beta_{ih}B')$. Let $\beta_h^*(P,d) \in (0,1)$ be the (unique) solution of $\beta = \Psi_h(P, d, \beta B', \beta B')$ and let $\beta_{ih}^* = \beta_h^*(P_{ih}, d_{ih}) \in (0,1)$ with $\beta_{ih} > \Psi_h(P_{ih}, d_{ih}, \beta_{ih}B', \beta_{ih}B')$ iff $\beta_{ih} > \beta_h^*(P_{ih}, d_{ih})$. We thus have $\alpha_{ih} = 1$ - respect. $\alpha_{ih} \in (0,1), \alpha_{ih} = 0$ - if $\beta_{ih} < \beta_{ih}^*$ - respect. $\beta_{ih} = \beta_{ih}^*, \beta_{ih} > \beta_{ih}^*$. Hence β_{ih}^* is the audit probability (for claims filed for severe accidents) above which type h individuals and repairers are deterred from colluding, when insurance has been purchased through distribution channel i.

1.4 Equilibrium fraud and audit

Let L_1 and L_2 be the expected repair costs, conditionally upon the occurrence of one or two accidents respectively, with $L_1 = (q_m + 2q_s)\ell$ and $L_2 = 2(q_m^2 + 2q_s^2 + 3q_mq_s)\ell$.[12] The expected

[11] $\Psi_h(P_{ih}, d_{ih}, G_{SP}, G_{NSP})$ is just the value of β_{ih} such that $Eu_{ih}^F = Eu_{ih}^N$.

[12] If one single accident occurs, it is minor with probability q_m and severe with probability q_s, with costs ℓ and 2ℓ, respectively . In the case of two accidents, both of them are minor with probability q_m^2 and cost 2ℓ, or

cost of claims may be written as:

$$C_{ih} = \overline{L} - (\pi_1 + 2\pi_2)d_{ih} + FC_{ih} + AC_{ih}, \tag{A-2}$$

where $\overline{L} = \pi_1 L_1 + \pi_2 L_2$ is the expected repair cost, FC_{ih} is the expected cost of fraudulent claims, and AC_{ih} is the expected audit cost. Thus $\overline{L} - (\pi_1 + 2\pi_2)d_{ih}$ is the share of the expected repair cost borne by the insurer, and $FC_{ih} + AC_{ih}$ is the total cost of fraud. Let us express FC_{ih} and AC_{ih} as functions of fraud and audit strategies. We have:

$$\begin{aligned} FC_{ih} &= \delta q_m \alpha_{ih}(\pi_1 + \pi_2)(1-\mu) \\ &\quad \times \{q_m \widehat{\pi}[(1-\beta_{ih})(d_{ih} + \mu v) - 2\beta_{ih}(\ell - d_{ih})] \\ &\quad -(1 - q_m\widehat{\pi})(\ell - d_{ih})\}, \end{aligned} \tag{A-3}$$

which may be read as follows. A policyholder who intends to renew his contract (which represents a fraction δ of all policyholders) may try to defraud if he has at least one accident, the first one being minor and in NSP: this case occurs with probability $q_m(\pi_1 + \pi_2)(1-\mu)$. He then postpones his claim with probability α_{ih}, and he will actually have the opportunity to defraud with probability $q_m\widehat{\pi}$. In that case, fraud will be detected with probability β_{ih}, and no insurance indemnity will be paid for the two minor claims, hence the gain $2(\ell - d_{ih})$ for the insurer. With probability $1-\beta_{ih}$, fraud is not detected and the additional cost to the insurer is $d_{ih} + v$ or d_{ih} if the fraudulent claim is filed in SP or NSP, i.e., with probability μ and $1-\mu$, respectively. If the policyholder does not have the opportunity to defraud (which occurs with probability $1 - q_m\widehat{\pi}$), he just loses the indemnity for the first claim $\ell - d_{ih}$.[13]

Furthermore, we have $AC_{ih} = N_{ih}c_i$, where N_{ih} is the number of audits per type h policyholder for distribution channel i. Audits are concentrated on the first claims that correspond to severe accidents. Policyholders have at least one accident, the first one being severe, with probability $q_s(\pi_1 + \pi_2)$. In addition, opportunistic policyholders who intend to renew their contract file a fraudulent claim with probability $q_m^2\alpha_{ih}\pi_2(1-\mu)$.[14] Severe accident claims are audited with probability β_{ih}, regardless of when the accidents are reported. Thus,

both are severe with probability q_s^2 and cost 4ℓ, or one is minor and the other one is severe with probability $2q_mq_s$ and cost 3ℓ.

[13] Keep in mind that fines B and B' are not part of the insurer's income.

[14] Indeed, the policyholder has two minor accidents, the first one in NSP with probability $q_m^2\pi_2(1-\mu)$. He does not file a claim immediately after the first accident with probability α_{ih}.

we have:

$$AC_{ih} = N_{ih}c_i = \beta_{ih}c_i[q_s(\pi_1 + \pi_2) + \delta q_m^2 \alpha_{ih}\pi_2(1-\mu)]. \tag{A-4}$$

The audit probability β_{ih} is chosen in $[0,1]$ by the insurer to minimize the expected cost of claims C_{ih}. We thus have $\beta_{ih} = 1$ - respect. $\beta_{ih} \in (0,1), \beta_{ih} = 0$ - if $\alpha_{ih} < \alpha^*_{ih}$ - respect. $\alpha_{ih} = \alpha^*_{ih}, \alpha_{ih} > \alpha^*_{ih}$- where $\alpha^*_{ih} = \alpha^*(d_{ih}, c_i)$, with

$$\alpha^*(d,c) \equiv \frac{q_s c(\pi_1 + \pi_2)}{\delta\pi_2 q_m^2 \mu(1-\mu)(2\ell - d + \mu v - c)}. \tag{A-5}$$

α^*_{ih} is the threshold fraud rate such that the insurer is incentivized to audit claims if and only if $\alpha_{ih} \geq \alpha^*_{ih}$. We have $\alpha^*_{ih} \in (0,1)$ if c_i is not too large, and we focus attention on this case in what follows.

At equilibrium, the decisions of the policyholder-repairer coalition and of the insurer should be mutual best responses. The equilibrium is in mixed strategies: insurers audit claims with a probability that makes the potential defrauder (here the policyholder-repairer coalition) indifferent between defrauding and not defrauding, and symmetrically, the fraud rate makes insurers indifferent between auditing and not auditing. This is stated in Proposition 1.

Proposition 1 *When insurers offer contract $(P_{i1}, d_{i1}), (P_{i2}, d_{i2})$ through $i \in \{D, A\}$, the equilibrium fraud rates and the equilibrium audit strategies are $\alpha_{ih} = \alpha^*(d_{ih}, c_i)$ and $\beta_{ih} = \beta^*_h(P_{ih}, d_{ih})$, respectively.*

Corollary 1 *For any distribution channel $i \in \{A, D\}$, we have $\alpha_{i2} > \alpha_{i1}$ iff $d_{i2} > d_{i1}$, i.e., the larger the deductible, the larger the fraud rate.*

Corollary 1 is a direct consequence of Proposition 1 because $\alpha^*(d,c)$ is increasing in d. The larger the deductible, the smaller the insurer's incentives to audit the claim, and thus the larger the minimal fraud rate that incentivizes the insurer to perform audits. In particular, everything else given (and in particular for a given distribution channel), the model predicts a larger fraud rate for deductible contracts than for full coverage contracts.

When $\alpha_{ih} = \alpha^*(d_{ih}, c_i)$, the expected cost of an insurance policy purchased by type h

individuals through channel i is

$$C_{ih} = \overline{L} - (\pi_1 + 2\pi_2)d_{ih} + k_0(d_{ih} + k_1)\alpha^*(d_{ih}, c_i),$$

where $k_0 = q_m(\pi_1 + \pi_2)(1 - \mu) \in (0, 1)$ and $k_1 = q_m\mu\widehat{\pi}v + \ell(1 - q_m\mu\widehat{\pi})$. Insurers price their contracts with the loading factor $\sigma > 0$. Thus, we have:

$$\begin{aligned} P_{ih} &= (1+\sigma)C_{ih} \\ &= (1+\sigma)[\overline{L} - (\pi_1 + 2\pi_2)d_{ih} + k_0(d_{ih} + k_1)\alpha^*(d_{ih}, c_i)], \end{aligned}$$

which may be written more compactly as $P_{ih} = \Phi(d_{ih}, \alpha^*(d_{i1h}, c_i))$,where

$$\Phi(d, \alpha) \equiv (1+\sigma)[\overline{L} - (\pi_1 + 2\pi_2)d + k_0(d + k_1)\alpha].$$

It is assumed that competition betwen insurers allows policyholders to extract all the surplus of the insurance contract. This surplus is independent from the individual's preferences between the two distribution channels, i.e., from the search costs tx and $t(1 - x)$ when the insurance seekers choose to purchase insurance from D or A, respectively. Thus, the equilibrium contract (P_{ih}, d_{ih}) maximizes $\overline{u}_h(P, d)$ subject to $P = \Phi(d, \alpha^*(d, c_i))$, and the equilibrium fraud rates are $\alpha_{ih} = \alpha^*(d_{ih}, c_i)$ for $h \in \{1, 2\}$,$i \in \{A, D\}$.

Proposition 2 *The optimal insurance contracts are such that $d_{i2} \geq d_{i1} \geq 0$, with $d_{i2} > d_{i1}$ if $d_{i1} > 0$ for $i = A$ or D.*

The extent of coverage is the result of a trade-off between the incentives to audit claims and the transaction costs, materialized by the fact that the fraud rate $\alpha^*(d, c)$ is increasing in d and by the loading factor σ, respectively. In the absence of transaction costs, overcoverage would be optimal.[15] We have excluded overcoverage so that full coverage would be optimal if there were no transaction costs. However, transaction costs reduce the optimal insurance coverage. Type 1 individuals are more risk averse than type 2 individuals, and thus Proposition 2 states that their deductible is smaller, as in the usual comparative statics of deductible contracts (see Schlesinger (2013)). The trade-off between increasing audit incentives and reducing transaction costs may tip in favor of positive deductibles for type

[15] See Boyer (2004).

2 and full coverage for type 1, and in that case deductible and no-deductible contracts are simultaneously offered at equilibrium.[16]

Proposition 3 *At equilibrium, we have* $\alpha_{D1} > \alpha_{A1}, \alpha_{D2} > \alpha_{A2}$*, that is, for both types of individuals the fraud rate is larger among insurance policies purchased through* D *than through* A*.*

Insurers need additional incentives to audit claims when insurance policies have been purchased through D than through A, because establishing the truth is more costly in the first case than in the second (i.e., $c_D > c_A$). These additional incentives emerge when the fraud rate is higher, which corresponds to the fact that $\alpha^*(d, c)$ is increasing with c, hence at equilibrium there is a higher fraud rate for D than for A. The proof of Proposition 3 shows that this basic intuition remains valid if we take into account the fact that optimal deductibles may differ between both cases (i.e., we may have $d_{Dh} \neq d_{Ah}$), which also affect incentives.

Finally, the market shares of D and A are defined by the threshold $x_h^* \in [0, 1]$ such that type h individuals located at $x \in [0, 1]$ choose $i = D$ if $x < x_h^*$, and they choose $i = A$ if $x > x_h^*$. We have $P_{ih} = \Phi(d_{ih}, \alpha_{ih})$ for $i \in \{D, A\}$. Hence

$$\overline{u}_h(\Phi(d_{Dh}, \alpha_{Dh}), d_{Dh}) - tx_h^* = \overline{u}_h(\Phi(d_{Ah}, \alpha_{Ah}), d_{Ah}) - t(1 - x_h^*),$$

and thus the market shares of D and A are characterized by

$$x_h^* = \frac{1}{2} + \frac{\overline{u}_h(\Phi(d_{Dh}, \alpha_{Dh}), d_{Dh}) - \overline{u}_h(\Phi(d_{Ah}, \alpha_{Ah}), d_{Ah})}{2t},$$

for $h = 1, 2$.

1.5 Proofs

Proof of Proposition 1

If $\beta_{ih} > \beta_{ih}^*$, then the optimal choice of the policyholder is $\alpha_{ih} = 0 < \alpha_{ih}^*$, which gives $\beta_{ih} = 0$ for the optimal choice of the insurer, hence a contradiction. Symmetrically, if

[16] See the illustrative example with mean-variance preferences below.

$\beta_{ih} < \beta^*_{ih}$, then the optimal choice of the policyholder is $\alpha_{ih} = 1 > \alpha^*_{ih}$, which gives $\beta_{ih} = 1$ for the optimal choice of the insurer, hence once again a contradiction. Thus we necessarily have $\beta_{ih} = \beta^*_{ih} \in (0,1)$ at equilibrium. $\beta_{ih} = \beta^*_{ih}$ is an optimal choice of the insurer if $\alpha_{ih} = \alpha^*_{ih}$. Symmetrically, $\alpha_{ih} = \alpha^*_{ih} \in (0,1)$ is an optimal choice of the policyholder if $\beta_{ih} = \beta^*_{ih}$. Thus $\alpha_{ih} = \alpha^*_{ih}, \beta_{ih} = \beta^*_{ih}$ is the unique equilibrium.

Proof of Proposition 2

Let the expected utility of type $h = 1, 2$ policyholders who purchase insurance through $i = A, D$ be written as $\widetilde{u}_{ih}(d) \equiv \overline{u}_h(\Phi(d, \alpha^*(d, c_i)), d)$, where $\Phi(.)$ and $\alpha^*(.)$ are defined by

$$
\begin{aligned}
\Phi(d,\alpha) &= (1+\sigma)[\overline{L} - (\pi_1 + 2\pi_2)d + k_0(d+k_1)\alpha], \\
\alpha^*(d,c) &= Kc(2\ell - d + v - c)^{-1},
\end{aligned}
$$

with $K \equiv q_s(\pi_1 + \pi_2)/\pi_2 q_m^2 \mu(1-\mu)$. Let $\widetilde{\Phi}_i(d) \equiv \Phi(d, \alpha^*(d, c_i))$. We have

$$
\begin{aligned}
\widetilde{\Phi}'_i(d) &= (1+\sigma)\left\{ -(\pi_1 + 2\pi_2) + k_0\alpha^*(d,c_i) + \frac{k_0 Kc(d+k_1)}{(2\ell - d + v - c_i)^2} \right\} \\
\widetilde{\Phi}''_i(d) &= 2Kck_0(1+\sigma)(2\ell - d + v - c_i)^{-3}(2\ell + v - c_i + k_1) > 0.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\widetilde{u}_{ih}(d) &= (1 - \pi_1 - \pi_2)u_h(w - \widetilde{\Phi}_i(d)) + \pi_1 u_h(w - \widetilde{\Phi}_i(d) - d - \varepsilon) \\
&\quad + \pi_2 u_h(w - \widetilde{\Phi}_i(d) - 2d - 2\varepsilon), \\
\widetilde{u}'_{ih}(d) &= -(1 - \pi_1 - \pi_2)u'_h(w - \widetilde{\Phi}_i(d))\widetilde{\Phi}'_i(d) \\
&\quad - \pi_1 u'_h(w - \widetilde{\Phi}_i(d) - d - \varepsilon)[1 + \widetilde{\Phi}'_i(d)] \\
&\quad - \pi_2 u'_h(w - \widetilde{\Phi}_i(d) - 2d - 2\varepsilon)[2 + \widetilde{\Phi}'_i(d)].
\end{aligned}
$$

Using $\widetilde{\Phi}''_i(d) > 0$ and $u''_h < 0$ shows that $\widetilde{u}_{ih}(d)$ is a concave function. Let d_{ih} be the optimal deductible for type h individuals, i.e., d_{ih} maximizes $\widetilde{u}_{ih}(d)$ with respect to $d \geq 0$. Assume first that $d_{i2} > 0$, which implies $\widetilde{u}'_{i2}(d_{i2}) = 0$ and $\widetilde{\Phi}'_i(d_{i2}) < 0$. We have

$$
\begin{aligned}
\widetilde{u}'_{i1}(d_{i2}) &= -(1 - \pi_1 - \pi_2)u'_1(w - \widetilde{\Phi}_i(d_{i2}))\widetilde{\Phi}'_i(d_{i2}) \\
&\quad - \pi_1 u'_1(w - \widetilde{\Phi}_i(d_{i2}) - d_{i2} - \varepsilon)[1 + \widetilde{\Phi}'_i(d_{i2})] \\
&\quad - \pi_2 u'_1(w - \widetilde{\Phi}_i(d_{i2}) - 2d_{i2} - 2\varepsilon)[2 + \widetilde{\Phi}'_i(d_{i2})].
\end{aligned}
$$

Since type 1 individuals are more risk averse than type 2 individuals, we know from Pratt (1964) that there exists a function $g : \mathbb{R} \longrightarrow \mathbb{R}$ such that $u_1(y) \equiv g(u_2(y))$, with $g' > 0$ and $g'' < 0$. This allows us to write

$$\begin{aligned}
\widetilde{u}'_{i1}(d_{i2}) &= -(1 - \pi_1 - \pi_2)g'(u_2(y_0))u'_2(y_0)\widetilde{\Phi}'_i(d_{i2}) \\
&\quad -\pi_1 g'(u_2(y_1))u'_2(y_1)(1 + \widetilde{\Phi}'_i(d_{i2})) \\
&\quad -\pi_2 g'(u_2(y_2))u'_1(y_2)(2 + \widetilde{\Phi}'_i(d_{i2})),
\end{aligned}$$

where $y_0 = w - \widetilde{\Phi}_i(d_{i2}), y_1 = w - \widetilde{\Phi}_i(d_{i2}) - d_{i2} - \varepsilon$ and $y_2 = w - \widetilde{\Phi}_i(d_{i2}) - 2d_{i2} - 2\varepsilon$, with $y_2 < y_1 < y_0$. Let us first consider the case where $1 + \widetilde{\Phi}'_i(d_{i2}) > 0$. Let $y^* \in (y_1, y_0)$. Using $g'' < 0$ and $u'_2 > 0$ yields

$$g'(u_2(y_0)) < g'(u_2(y^*)) < g'(u_2(y_1)) < g'(u_2(y_2)).$$

Using $\widetilde{\Phi}'_i(d_{i2}) < 0 < 1 + \widetilde{\Phi}'_i(d_{i2})$ then gives

$$\widetilde{u}'_{i1}(d_{i2}) < g'(u_2(y^*))\widetilde{u}'_{i2}(d_{i2}) = 0,$$

which implies $d_{i1} < d_{i2}$ because of the concavity of $\widetilde{u}_{i1}(d)$. Similarly, when $1 + \widetilde{\Phi}'_i(d_{i2}) < 0 < 2 + \widetilde{\Phi}'_i(d_{i2})$, we let $y^* \in (y_2, y_1)$ and a similar argument also yields $d_{i1} < d_{i2}$. Similarly, if $d_{i2} = 0$, we have $\widetilde{u}'_{i1}(0) \leq 0$, and the same argument gives $\widetilde{u}'_{i2}(0) < 0$ and thus $d_{i1} = 0$.

Example with mean-variance preferences

The case $d_{i1} = 0, d_{i2} > 0$ can be conveniently illustrated by a mean-variance example. Assume that $u_1(w_f)$ and $u_2(w_f)$ are quadratic, so that we may write

$$u_h(w_f) = E(w_f) - \eta_h Var(w_f),$$

with $\eta_1 > \eta_2$. When insurance is purchased through distribution channel i, we have

$$\begin{aligned}
E(w_f) &= w - \widetilde{\Phi}_i(d) - (\pi_1 + 2\pi_2)(d + \varepsilon), \\
Var(w_f) &= I(d + \varepsilon)^2,
\end{aligned}$$

with $I = \pi_1(1-\pi_1) + 4\pi_2(1-\pi_1-\pi_2) > 0$. We have $d_{ih} > 0$ iff

$$-\widetilde{\Phi}_i'(0) - (\pi_1 + 2\pi_2) - 2\eta_h I\varepsilon^2 > 0,$$

and thus we have $d_{i1} = 0, d_{i2} > 0$ if $\eta_2 < \eta_i^* < \eta_1$, where

$$\eta_i^* = \frac{1}{2I\varepsilon^2} \times \left[\sigma(\pi_1 + 2\pi_2) - \frac{k_0 K c_i (1+\sigma)(2\ell + v - c_i + k_1)}{(2\ell + v - c_i)^2}\right]$$

Proof of Proposition 3

Let us write:

$$\overline{u}_h(\Phi(d,\alpha), d) \equiv \Gamma_h(d,\alpha),$$
$$\alpha^*(d,c) = Kc(2\ell - d + v - c)^{-1},$$

with $K \equiv q_s(\pi_1 + \pi_2)/\pi_2 q_m^2 \mu(1-\mu)$. Assume $\overline{u}_h(\Phi(d,\alpha^*(d,c)),d) = \Gamma_h(d,\alpha^*(d,c))$ is maximized w.r.t. d at $d = \widehat{d}_h(c)$ with fraud rate $\widehat{\alpha}_h(c) \equiv \alpha^*(\widehat{d}_h(c), c)$. Thus, we have $\alpha_{ih} = \widehat{\alpha}_h(c_i) \equiv \alpha^*(\widehat{d}_h(c_i), c_i)$ for $i \in \{A, D\}$. The first-order and second-order optimality conditions for this maximization are respectively written as:

$$F_h \equiv \Gamma'_{hd} + \Gamma'_{h\alpha}\frac{\partial\alpha^*}{\partial d} = 0, \tag{A-6}$$
$$S_h \equiv \Gamma''_{hd^2} + \Gamma''_{hd\alpha}\frac{\partial\alpha^*}{\partial d} + \Gamma''_{h\alpha^2}\left(\frac{\partial\alpha^*}{\partial d}\right)^2 + \Gamma'_{h\alpha}\frac{\partial^2\alpha^*}{\partial d^2} < 0, \tag{A-7}$$

where $\Gamma'_{hd}, \Gamma'_{h\alpha}, \Gamma''_{hd^2}, \Gamma''_{hd\alpha}, \Gamma''_{h\alpha^2}$ denote first and second derivatives of Γ_h and all functions are evaluted at $d = \widehat{d}_h(c)$. Differentiating (A-6) gives $\widehat{d}'_h(c) = -F'_{hc}/S_h$ where:

$$F'_{hc} = \frac{\partial F_h}{\partial c} = \Gamma''_{hd\alpha}\frac{\partial\alpha^*}{\partial c} + \Gamma''_{h\alpha^2}\frac{\partial\alpha^*}{\partial c}\frac{\partial\alpha^*}{\partial d} + \Gamma'_{h\alpha}\frac{\partial^2\alpha^*}{\partial d\partial c}.$$

After simplication we get:

$$\begin{aligned}\widehat{\alpha}'_h(c) &= \frac{\partial\alpha^*}{\partial d}\widehat{d}'_h(c) + \frac{\partial\alpha^*}{\partial c}\\ &= (1/S_h)\left\{\Gamma'_{h\alpha}\left[\frac{\partial^2\alpha^*}{\partial d^2}\frac{\partial\alpha^*}{\partial c} - \frac{\partial^2\alpha^*}{\partial d\partial c}\frac{\partial\alpha^*}{\partial d}\right] + \Gamma''_{hd^2}\frac{\partial\alpha^*}{\partial c}\right\},\end{aligned} \tag{A-8}$$

with

$$\frac{\partial^2\alpha^*}{\partial d^2}\frac{\partial\alpha^*}{\partial c} - \frac{\partial^2\alpha^*}{\partial d\partial c}\frac{\partial\alpha^*}{\partial d} = K^2 c(2\ell - d - c)^{-4} > 0,$$

and

$$\begin{aligned}\Gamma_h(d,\alpha) &= (1-\pi_1-\pi_2)u_h(w-\Phi(d,\alpha)) \\ &+\pi_1 u_h(w-\Phi(d,\alpha)-d) + \pi_2 u_h(w-\Phi(d,\alpha)-2d).\end{aligned}$$

$\Phi(d,\alpha)$ is linear in d, and thus $\Gamma_h(d,\alpha)$ is concave in d, which implies $\Gamma''_{hd^2} < 0$. We also have $\Gamma'_{h\alpha} = (\partial\overline{u}_h/\partial P) \times (\partial\Phi/\partial\alpha) < 0$. Using (A-8) and $\Gamma'_{h\alpha} < 0, \Gamma''_{hd^2} < 0, \partial\alpha^*/\partial c > 0$ then yields $\widehat{\alpha}'_h(c) > 0$. Thus, we have $\alpha_{Dh} = \widehat{\alpha}_h(c_D) > \widehat{\alpha}_h(c_A) = \alpha_{A1h}$.

1.6 Fraud and bargaining power

We may adapt the previous model in order to show how the bargaining power of the policyholder-repairer coalition affects the scale of fraud. As in Section 3 of the paper, the bargaining power of the colluders is taken into account by assuming that the defrauders will not be punished with probability $\xi_i \in (0,1)$, with $i = D$ or A. Intuitively, the insurance agent is incentivized to stand up for its customer (and possibly also for the repairer in the case of a DOA that owns the repair shop), and it may threaten the insurer to redirect its customers toward another insurer. This may deter the insurer from enforcing the penalty. A larger bargaining power for D than for A corresponds to $\xi_D > \xi_A$. Thus, if the colluders are spotted (which occurs if the claim is audited), then with probability $1-\xi_i$ the penalties are enforced (no indemnity is paid by the insurer and the colluders pay the fines B and B', respectively), and with probability ξ_i the insurer interprets the fraud as an involuntary error, i.e., the policyholder receives the total cumulated contractual indemnity $2(\ell - d_{ih})$ and no fines are paid. Under these assumptions, a type h policyholder with two minor accidents and a repairer are willing to defraud (with the policyholder making a side-payment G to the repairer on a take-it-or-leave-it basis) if

$$Eu^F_{ih} \geq Eu^N_{ih}, \tag{A-9}$$

and

$$G - \beta_{ih}(1-\xi_i)B' \geq 0, \tag{A-10}$$

respectively, with

$$\begin{aligned} Eu_{ih}^F &= q_m\mu\widehat{\pi}[(1-\beta_{ih})u_h(w-P_{ih}-d_{ih}-2\varepsilon+v-G) \\ &\quad +\beta_{ih}(1-\xi_i)u_h(w-P_{ih}-2\ell-2\varepsilon-G-B)+\beta_{ih}\xi_i u_h(w-P_{ih}-2d_{ih}-G)] \\ &\quad +(1-q_m\mu\widehat{\pi})[\widehat{\pi}u_h(w-P_{ih}-\ell-d_{ih}-2\varepsilon)+(1-\widehat{\pi})u_h(w-P_{ih}-\ell-\varepsilon)]. \end{aligned}$$

Fraud is deterred (i.e., $\alpha_{ih}=0$) if $\beta_{ih} > \beta_{ih}^{**}$ where $\beta_{ih}^{**}(P_{i1}, d_{ih}, \xi_i)$ is the value of β_{ih} such that

$$Eu_{ih}^F = Eu_{ih}^N \text{ with } G = \beta_{ih}(1-\xi_i)B'.$$

Since defrauders who are caught are not punished with probability ξ_i, the expected actuarial cost of a deductible insurance policy is now written as

$$\begin{aligned} FC_{ih} &= \delta q_m \alpha_{ih}(\pi_1+\pi_2)(1-\mu) \\ &\quad \times\{q_m\mu\widehat{\pi}[(1-\beta_{ih})(d_{ih}+v)-2\beta_{ih}(1-\xi_i)(\ell-d_{ih})] \\ &\quad -(1-q_m\mu\widehat{\pi})(\ell-d_{ih})\}. \end{aligned} \tag{A-11}$$

The equilibrium audit and fraud strategies are $\alpha_{ih}=\alpha^{**}(d_{ih}, c_i, \xi_i)$ and $\beta_{ih}=\beta_{ih}^{**}(P_{ih}, d_{ih}, \xi_i)$, with

$$\alpha_{ih}^{**}(d_{ih}, c_i, \xi_i) \equiv \frac{q_s c_i(\pi_1+\pi_2)}{\pi_2 q_m^2 \mu(1-\mu)[(1-\xi_i)(2\ell-d_{ih})+\xi_i d+v-c_i)}, \tag{A-12}$$

which can be interpreted in the same way as (A-5). The equilibrium contract (P_{ih}, d_{ih}) maximizes $\overline{u}_h(P,d)$ subject to $P=\Phi(d, \alpha^{**}(d, c_i, \xi_i))$, and the equilibrium fraud rates are $\alpha_{ih}=\alpha^{**}(d_{ih}, c_i, \xi_i)$ for $i=A$ or D. In the same way as in the proof of Proposition 3, we can then show that $\alpha_{Dh} > \alpha_{Ah}$ if $c_D = c_A$ and $\xi_D > \xi_A$. To establish this result, we make the additional assumption $\xi_D < 1/2$.[17] The definition of $\Phi(d,\alpha)$ is unchanged, and we still denote $\Gamma_h(d,\alpha) \equiv \overline{u}_h(\Phi(d,\alpha),d)$.

$\Gamma_h(d,\alpha^{**}(d,c,\xi))$ is maximized w.r.t. d at $d = \widetilde{d}_h(c,\xi)$, with fraud rate $\widetilde{\alpha}_h(c,\xi) \equiv \alpha^{**}(\widetilde{d}_h(c,\xi),c,\xi)$. The equilibrium fraud rates are $\alpha_{ih} = \widetilde{\alpha}_h(c_i,\xi_i) \equiv \alpha^{**}(\widetilde{d}_h(c_i,\xi_i),c_i,\xi_i)$ for $i \in \{A,D\}$. We have (similarly to the proof of Proposition 3, with an unchanged definition for S_h):

$$\begin{aligned}
\frac{\partial\widetilde{\alpha}(c,\xi)}{\partial\xi} &= \frac{\partial\alpha^{**}}{\partial d}\frac{\partial\widetilde{d}_h(c,\xi)}{\partial\xi} + \frac{\partial\alpha^{**}}{\partial\xi} \\
&= (1/S_h)\left\{\Gamma'_{h\alpha}\left[\frac{\partial^2\alpha^{**}}{\partial d^2}\frac{\partial\alpha^{**}}{\partial\xi} - \frac{\partial^2\alpha^{**}}{\partial d\partial\xi}\frac{\partial\alpha^{**}}{\partial d}\right] + \Gamma''_{hd^2}\frac{\partial\alpha^{**}}{\partial\xi}\right\}. \qquad \text{(A-13)}
\end{aligned}$$

We have $\alpha^{**}(d,c,\xi) = Kc[(1-\xi)(2\ell-d)+\xi d+v-c]^{-1}$, and thus

$$\begin{aligned}
&\frac{\partial\alpha^{**}}{\partial d} = Kc(1-2\xi)[(1-\xi)(2\ell-d)+\xi d+v-c]^{-2}, \\
&\frac{\partial\alpha^{**}}{\partial\xi} = 2Kc(\ell-d)[(1-\xi)(2\ell-d)+\xi d+v-c]^{-2}, \\
&\frac{\partial^2\alpha^{**}}{\partial d^2} = -2Kc(1-2\xi)^2[(1-\xi)(2\ell-d)+\xi d+v-c]^{-3}, \\
&\frac{\partial^2\alpha^{**}}{\partial d\partial\xi} = -2Kc[(1-\xi)(2\ell-d)+\xi d+v-c]^{-2} \\
&\qquad - 4Kc(1-2\xi)(\ell-d)[(1-\xi)(2\ell-d)+\xi d+v-c]^{-3}.
\end{aligned}$$

Hence,

$$\frac{\partial^2\alpha^{**}}{\partial d^2}\frac{\partial\alpha^{**}}{\partial\xi} - \frac{\partial^2\alpha^{**}}{\partial d\partial\xi}\frac{\partial\alpha^{**}}{\partial d} = 2K^2c^2(1-2\xi)[(1-\xi)(2\ell-d)+\xi d-c]^{-4} > 0.$$

Using $S_h < 0, \Gamma'_{h\alpha} < 0, \Gamma''_{hd^2} < 0, \partial\alpha^{**}/\partial\xi > 0$, (A-13) yields $\partial\widetilde{\alpha}(c,\xi)/\partial\xi > 0$. Thus, we have $\alpha_{Dh} = \widetilde{\alpha}(c,\xi_D) > \widetilde{\alpha}(c,\xi_A) = \alpha_{A1}$ when $c_D = c_A = c$ and $\xi_D > \xi_A$.

2 Complements to the empirical analysis

[17] For a given fraud rate α_{ih}, the decrease in actuarial cost $\mathrm{d}C_{ih} < 0$ induced by a small increase in the audit probability $\mathrm{d}\beta_{ih} > 0$ is $\mathrm{d}C_{ih} = -\eta q_m^2\alpha_{ih}[d_{ih} + 2(1-\xi_i)(\ell - d_{ih})]\mathrm{d}\beta_{ih}$. We consider the case where the decrease in cost is larger when the deductible is lower, which requires $\xi_i < 1/2$.

2.1 Validating the IV approach

Table 8 provides three tests that confirm the properness of our IV approach. In the two first stage of LPMs, the Anderson-Rubin test and the J test do not reject the null hypothesis of the exogenous instrumental variable and the null hypothesis of no over identified instruments. The Durbin-Wu-Hausman test rejects the null hypothesis of no endogeneity. We also find the consistent results obtained in the 2SLS approach that policyholders living in areas with high average income or low percentage of highly educated people significantly tend to choose policies with deductibles and not to renew their contract.

2.2 Testing Hypothesis 1 through the Chiappori-Salanié (2000) approach

Chiappori and Salanié (2000) use a pair of Probit regressions to explain the probability of filing a claim and the probability of choosing partial coverage, and they appraise the conditional dependence between these two variables by submitting the residuals of the two regressions to a W test. Similarly, we have run two sets of pairwise Probit regressions, respectively with $SG1$ and SC, and with $SG2$ and SC as dependent variables. The W statistics, calculated with the residuals of each pairwise regressions, are significantly different from 0 at the 1% threshold. In each case, we have also calculated the correlation coefficient of these residuals: both are positive and significantly different from 0 at the 1% level, which confirms the validity of Hypothesis 1.[18]

2.3 Additional test for Hypothesis 2

The regression that explains the value of the claims has also been performed over the whole sample (not only the SG group) by including dummies $SG1_i, SG2_i, SC_i$, $first_i$, and their double and triple interaction terms in the explanatory variables. Furthermore, in order to be able to identify fraud (as defined above, that is claims manipulation) and the premium recouping behavior, we also include RG_i, and the double and triple interaction variables

[18] Computing the W statistic with the residuals of the regressions for $SG1$ and SC yields $W = 201.76$, which is significantly different from 0 at the 1% level. The correlation coefficient between the residuals of these regressions is $\rho = 0.003611$, and it is also significantly different from 0 at the 1% level. Likewise, using the residuals from the regressions for $SG2$ and SC gives $W = 257.99$ and $\rho = 0.03221$, and these statistics are significantly different from 0 at the 1% level. The full regression results are available from the authors upon request.

$RG_i * SC_i$, and $RG_i * SC_i * first_i$ among explanatory variables.

$$\begin{aligned} clmamt_i &= \alpha_{s1}SG1_i + \alpha_{s2}SG2_i + \alpha_{RG}RG_i + \alpha_c SC_i + \alpha_f first_i \\ &+\alpha_{cf}SC_i * first_i + \alpha_{sf1}SG1_i * first_i + \alpha_{sf2}SG2_i * first_i \\ &+\alpha_{Rf}RG_i * first_i + \alpha_{sc1}SG1_i * SC_i + \alpha_{sc2}SG2_i * SC_i \\ &+\alpha_{Rc}RG_i * SC_i + \alpha_{scf1}SG1_i * SC_i * first_i \\ &+\alpha_{scf2}SG2_i * SC_i * first_i + \alpha_{Rcf}RG_i * SC_i * first_i + \alpha X_i. \end{aligned} \tag{A-6}$$

Performing this regression among the 69,082 claims filed by the members of the research sample gives $\widehat{\alpha}_{scf1} = 96.7$ with $p-$value <0.012, $\widehat{\alpha}_{scf2} = 235.1$ with $p-$value <0.0001, and $\widehat{\alpha}_{Rcf} = -97.3$ with $p-$value <0.086.[19] The inequalities $\widehat{\alpha}_{scf2} > \widehat{\alpha}_{scf1} > 0$ once again validate Hypothesis 2. Symmetrically, $\widehat{\alpha}_{Rcf} < 0$ confirms that members of the RG group tend to file small claims at the end of the policy year, when they have not filed any claim during the previous months.

2.4 Taking adverse selection into account

In a setting with adverse selection, past and future claim experiences may be linked, but man-made claim manipulation should reduce the predictive power of this link. To check if this is actually the case, we have used the 2010 data to run two Probit regressions that estimate the probability of filing a claim either in any month of 2011 or in the suspicious period of 2011, respectively. The regressions were run separately for the suspicious and non-suspicious groups.[20] Observing the policyholders' 2011 claim records allows us to calculate the prediction error for the claims filed in all of 2011 and for the claims filed in the suspicious period of 2011. In a second stage, we use a t-test to evaluate whether this prediction error is smaller for the claims filed over the whole year than for those filed in the suspicious period.[21] Panel A of Table 5 confirms that this is the case, at the same time for both the suspicious

[19] The full estimated results of regressions (11) and (A-6) are available from the authors upon request.

[20] In other words, these Probit regressions regress clm_i and SC_i, respectively, on the explanatory variables included in the vector of observable variables X_i.

[21] The prediction error is the absolute value of the difference between the estimated probability of filing a claim and the dummy equal to 1 if the individual has filed a claim in 2011 and 0 otherwise. We calculate the difference between the prediction errors over the whole 2011 year and over the suspicious period, and we test whether this difference is negative.

and non-suspicious groups. Furthermore, the difference of the prediction error is significantly different and larger in absolute value in the suspicious groups, especially in $SG2$, than in the non-suspicious group. This confirms the manipulation of claims, beyond any possible hidden information about policyholders' risk types.

Secondly, we know that adverse selection may lead to a positive correlation between the contract coverage and the probability of filing claims, but it does not induce any particular timing for claims such as the one on which we are focusing. Panel B of Table 5 provides the hazard rate in the suspicious groups $SG1$ and $SG2$, and in the non-suspicious group. In $SG1$ and $SG2$, the hazard rates are significantly higher in the last policy month than in the other months, and these last month hazard rates are significantly higher than in the non-suspicious group, which confirms that claim manipulation does occur. The fact that the last month hazard rate is even larger for $SG2$ than for $SG1$ confirms that our observations cannot be attributed to adverse selection.

2.5 Testing Hypothesis 1 for type B contracts only

Table 6 reports the results of two-stage regressions by limiting our sample to type B contracts.[22] The results are consistent with those of Table 3, which confirms the validity of Hypothesis 1.

2.6 Additional test for Hypothesis 3

It is legitimate to ask whether the higher expected cost of claims in the DOA channel simply reflects the fact that, on average, the individuals who take out insurance from DOAs have higher risks, rather than fraudulent behaviors. This issue may be clarified by estimating the claim amount using the following OLS regression:

$$claimamt_i = \alpha_0 + \alpha_D D_i + \alpha_A A_i + \alpha_B B_i + \alpha_{DA} D_i * A_i + \alpha_{DB} D_i * B_i + \alpha X_i + \varepsilon_i.$$

$claimamt_i$ is the claim amount which is estimated in thousand US dollars. A_i and B_i are dummies for type A or B contracts, respectively, with type C contract as counterpart,

[22] In other words, in this test, the suspicious group includes the policyholders (from the $SG1$ group) with a no-deductible type B contract that has been renewed at the end of the policy year, and the policyholders (from the $SG2$ group) with a deductible type B contract that has been renewed at the end of the policy year, while the control group contains the other policyholders with a type B contract that has not been renewed.

and X_i includes the underwriting and pricing variables as in the previous regressions. The estimated results are in Table 7. Type A and B contracts are associated with claim costs that are significantly larger than for type C contracts. The estimated coefficient of D_i is negative, but it is not significantly different from 0. Likewise, the estimated coefficients of interaction terms $D_i * A_i$, and $D_i * B_i$ are not significantly different from 0. In other words, the policyholders of the DOA channel do not have higher claim costs than others, whatever their contract. In other words, the increase in claim costs is not an intrinsic characteristic of the distribution channel: it reflects the fraudulent behaviors of some policyholders (the suspicious groups) who may take advantage of the manipulation of claims, and this behavior is facilitated by DOAs.

References

Boyer, M. (2004), "Overcompensation as a partial solution to commitment and renegotiation problerms: the case of ex post moral hazard", *Journal of Risk and Insurance,* 71(4): 559-582.

Chiappori, P.A. and Salanié, B. (2000), "Testing for asymmetric information in insurance markets", *Journal of Political Economy*, 108(1): 56–78.

Li, C-S., C-C. Liu and J-H. Yeh (2007), "The incentive effects of increasing per-claim deductible contracts in automobile insurance", *Journal of Risk and Insurance*, 74(2): 441-459.

Pratt, J. (1964), "Risk aversion in the small and in the large", *Econometrica*, 32: 122-136.

Schlesinger, H. (2013), "The theory of insurance demand", *Handbook of Insurance*, 2nd Edition, G. Dionne (Ed), Springer, 167-184.

Table 5: Additional evidence of fraud

	***SG*1**	***SG*2**	**non-*SG***
	Panel A: Predicted errors		
filing a claim	0.1127	0.1999	0.0832
filing *SC*	0.4328	0.5493	0.2605
t test	-138.3376(<0.0001)	-200 (<0.0001)	-36.6563(<0.0001)
	Panel B: Baseline hazard in each policy month		
1st month	0.0220	0.0253	0.0749
2nd month	0.0132	0.0157	0.0655
3rd month	0.0184	0.0158	0.0661
4th month	0.0126	0.0140	0.0638
5th month	0.0128	0.0154	0.0651
6th month	0.0136	0.0148	0.0647
7th month	0.0117	0.0148	0.0645
8th month	0.0151	0.0127	0.0630
9th month	0.0179	0.0129	0.0634
10th month	0.0200	0.0198	0.0699
11th month	0.0613	0.2210	0.1240
12th month	0.5401	0.6212	0.2633

Table 6: Empirical evidence of fraud - Focus on type B contracts

	First stage (bivariate Probit)		Second stage	
	SG	***deduct***	***2SLS-Probit***	***DGV-Probit***
constant	7.2958***	-1.7120***	-2.2382***	-2.1955***
Pr(*SG*1)			0.3556**	0.2596*
Pr(*SG2*)			0.4711***	0.3820**
SG1				1.0101***
SG2				1.0960***
income	-6.86E-06**	2.54E-05***		
edu	0.5871*	-1.7021**		
RG	-7.1345***	0.2422***	0.2747***	0.2155***
female	0.0118	-0.1863***	0.0071	0.0088
age2530	-0.4080**	0.1530	0.2421	0.2353
age3060	-0.3893**	-0.0214	0.2656	0.2602
ageabv60	-0.5109***	-0.0157	0.1796	0.1701
carage0	-0.5388***	-0.2287***	0.9048***	0.8933***
carage1	-0.1891***	-0.1593***	0.0917**	0.0875**
carage2	-0.1181**	-0.0785*	0.0167	0.0142
carage3	-0.0239	-0.0780*	-0.0226	-0.0232
carage4	-0.0552	-0.0556	-0.0226	-0.0230
veh_m	0.0155	0.0548*	-0.0559**	-0.0584**
veh_l	0.0911***	0.1027***	-0.1678***	-0.1721***
sedan	0.1076**	0.1369***	-0.0970**	-0.0983**
logprem	0.0707***	-0.2086***	0.1929***	0.2060***
bonus	-0.1377***	0.2890***	-0.5338***	-0.5580***
Pseudo R^2	0.1095		0.1062	0.4872

Same notes as in Table 3

Table 7: Comparing the risk between the policyholders from DOA and other distribution channels

Variables	Est. Ceoff.	P value
Intercept	-3.5765	<0.0001
D	0.0033	0.9200
A	0.3276	0.0020
B	0.5847	<0.0001
D*A	0.0513	0.1344
D*B	0.0580	0.1330
female	-0.0371	0.0370
age2530	0.1056	0.3930
age3060	-0.1425	0.2330
ageabv60	-0.1401	0.2530
carage0	0.1911	0.0000
carage1	0.0963	0.0010
carage2	0.0658	0.0380
carage3	0.0761	0.0150
carage4	0.0487	0.1550
veh_m	-0.0122	0.5430
veh_l	0.1944	<0.0001
sedan	0.2318	<0.0001
logprem	0.5236	<0.0001
bonus	0.4209	<0.0001

Note

In the above regression, we have also controlled the brand of the car. The results are not reported for confidentiality reasons.

Table 8: Empirical results from 2SLS-LPM

	2SLS-LPM			
	SG	***SC***	***deduct***	***SC***
constant	-3.0085***	-5.7209***	-0.3407	-4.6759***
Pr(*SG)*		0.3670***		
Pr(*deduct*)				0.2223***
income	-3.57E-06***		8.93E-06***	
edu	0.0839*		-0.3051**	
RG	-0.4447***	0.2212*	0.0297***	0.3838***
female	0.1222***	0.1185	-0.0051**	0.0710***
age2530	0.2590	0.5613***	0.0386	0.4713***
age3060	0.3809	0.6942***	0.0335	0.5551***
ageabv60	0.2970	0.5418**	0.0338	0.4324***
carage0	0.1501***	0.7874***	0.0446***	0.8329***
carage1	0.0789***	0.0062	0.0265***	0.0302
carage2	0.0591*	-0.0214	0.0161***	-0.0029
carage3	0.0280*	-0.0272	0.0072**	-0.0188
carage4	0.0204	-0.0263	0.0057	-0.0196
veh_m	0.1394	-0.1786**	0.0011	-0.1298***
veh_l	0.2793***	-0.4421***	0.0134***	-0.3394***
sedan	0.1395***	-0.2386***	0.0053	-0.1861***
logprem	0.5736***	0.8029**	0.0522**	0.5818***
bonus	-0.7123***	-1.2937***	-0.0490***	-1.0154***
	p-value		*p*-value	
J test	0.5665		0.5406	
Anderson-Rubin test	0.6981		0.7055	
Durbin-Wu-Hausman test	0.0576		0.0279	