Collusion between Retailers and Customers: The Case of Insurance Fraud in Taiwan

Pierre Picard* Jennifer Wang† Kili C. Wang‡

August 7th, 2023

Abstract

We analyze how insurance distribution channels may affect fraud through claim manipulation, when car repairers may collude with policyholders. We focus attention on the Taiwan automobile insurance market with a database provided by two large Taiwanese automobile insurers. The theoretical underpinning of our analysis is provided by a model of claims fraud with collusion and audit. Our econometric analysis confirms the evidence of fraud through the postponing of claims to the end of the policy year, possibly by filing a single claim for several events. It highlights the role of car dealer agencies in the fraud process, and its change from 2010 to 2018.

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

†Department of Risk Management and Insurance, National Chenghi University, Taiwan. Email: jenwang@nccu.edu.tw.

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

1 Introduction

Vertical relationships frequently involve the outsourcing of services from upstream firms to downstream retailers. This may be at the origin of agency costs, associated with the discretion in the way retailers do their job. Such agency 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. In such cases, dealers sell not only cars, but also automobile insurance to their clients, and furthermore they own an auto repair shop. Understandably, this multi-faceted activity and the long-term connection with car owners favor the creation of a mutually advantageous policyholder-DOA alliance. Concerning fraud itself, we will focus attention on two misbehaviors in the Taiwanese car insurance market. Firstly, the fact that policyholders may file small false claims by the end of the policy year if they have not receive any indemnity previously, a behavior highlighted by Li et al. (2013).

[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 by retailers are at the origin of a double moral hazard problem in warranty management.

Recouping a part of the insurance premium paid to presumably unfair insurers may be the psychological motivation behind this behavior. Secondly, postponing claims to the last month of the policy year and, when possible, merging two losses in a single claim. Deductibles and the bonus-malus mechanism are the underlying reasons for this second type of misbehavior. Disentangling these two types of fraud will be one of our main challenges in what follows.

An insurance market model yields the theoretical underpinnings of our analysis. The model focuses on the strategic interaction between, on the 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 the 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 bargaining power of DOAs may allow them to deter insurers from enforcing penalties. This suggests that there are potentially two reasons for which DOAs may facilitate insurance fraud: firstly, it may be hard for insurers to establish the truth because of the risk of collusion between DOAs and policyholders, and secondly, the bargaining power of DOAs may allow them not to be penalized when fraud is detected. The outcome is a higher fraud rate when insurance is distributed by DOAs than through other channels. As we will see below, this is reinforced in the case of deductible contracts, because deductibles increase the gain that policyholders obtain from fraud, and weaken the insurers' incentives to monitor claims.

Our empirical analysis draws on a database obtained from two large insurance companies in Taiwan. One of them, company 1, provided information on the policyholders who have filed an automobile claim in 2010 or 2018, and the other one, company 2, provided information on the policyholders who have filed an automobile claim in 2010. Company 1 relied heavily on DOAs to sell policies, although the market share of this distribution channel strongly decreased from 2010 to 2018, while company 2 never used DOAs.Starting with year 2010, our results confirm that there was more fraudulent claim manipulation when insurance policies were sold through DOAs than through other distribution channels, and also that deductibles stimulated fraud.[2] We also show that the causal mechanisms on which we focus (i.e., postponing claims, and possibly filing one claim for several accidents) were related to the bonus-malus system in force in Taiwan, and also to incentives 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, in 2010, the intertemporal pattern of claims was consistent with policyholder's fraudulent behavior favored by DOAs, after controlling for other explanations, including moral hazard and the money recouping behavior highlighted by Li et al. (2013).

However, things have changed dramatically from 2010 to 2018: DOAs were used less frequently by insurers, with presumably a lower bargaining power at the claim settlement stage. In other words, in 2018 it was more difficult for DOAs to collude with their customers at the expense of the insurer. As will be shown, the role of DOAs as facilitators of insurance fraud through claim manipulation vanished in 2018, in accordance with the decrease in their bargaining power.

The paper is organized as follows. Section 2 provides further motivation for our analy-

[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 claims 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 any fraud they are victims of. The preferred approach consists in 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 opportunistic theft insurance fraud by analysing the claim pattern in areas hit by a typhoon.

sis. We introduce some factual observations that should convince the reader that there is claim manipulation in the Taiwanese car insurance market, and we describe regular fraud patterns. Section 3 develops a simple theoretical model of insurance fraud that shows how these patterns are linked to specific features of insurance contracts, particularly per-claim deductibles, and to the insurance distribution channel.[4] Section 4 describes the data in more detail, it presents our econometric approach, and discusses our results about claim manipulation. We particularly highlight the changes in the fraud pattern and in the role of DOAs from 2010 to 2018. Section 5 concludes.

2 Factual background

Our investigation will be based on information yielded by two large Taiwanese insurers, refered to as companies 1 and 2, about their automobile policyholders and their claims in 2010 and 2018. In 2010, company 1 sold approximately 37% of its automobile policies through DOAs, and this share dropped to about 20% in 2018. On the contrary, company 2 never sold insurance through the DOA channel.

Insurance agents, be they DOAs or standard agents, are in charge of handling claims. This frequently involves some bargaining between the insurer whose objective is to minimize the cost of claims, and the agents, who may favour their customers, particularly when they receive sales-based commissions. In this bargaining process, DOAs take advantage of the size of their activity, and of the fact that they own the list of their customers. In particular, an insurer who discovers a claim manipulation by a DOA may be reluctant to take retaliatory actions because of this strategic advantage of DOAs, who could switch to another insurer.[5] In the case of company 1, the bargaining power of DOAs is expected

[4] This section may be skipped by readers who are mostly interested in the empirical analysis of insurance fraud.

[5] On average, Taiwanese DOAs sell more policies than other agents: three times more on average, and much more for the largest DOAs. They are independent agents, and, as emphasized by Mayers and Smith (1981), this status gives them more discretion in claim administration (e.g. they may intercede

to have decreased from 2010 to 2018, because this insurer has dramatically reduced its dependence on DOAs. The specificity of DOAs has also an informational dimension, related to the fact that they work in partnership with car repairers, both being sheltered by car dealers. This multifaceted agency relationship creates 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 specific forms of insurance contract manipulation 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 distribution of claims and their average cost (in hundred US dollars) in 2010 over the twelve policy months, with a striking concentration of claims and a slight decrease in the claim cost in the last months of the policy year. Li et al. (2013) interpret this phenomena as a "premium recouping effect": some policyholders without accident during the previous months would tend to file small false claims during the last month of the policy year to express their feeling that they have been unfairly treated by the insurance company.

Figure 1

Some information about insurance contracts is useful for what follows. There are three different types of automobile physical damage insurance contracts in Taiwan: types A, B and C. Types 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 the damages incurred in a collision involving two or more vehicles. Contracts also differ in terms of

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 company 1 than other insurance agents, with continuation rates (i.e. the fraction of customers who continue purchasing insurance from the same insurer year on year) 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.

indemnity: Type A contracts offer low coverage with a deductible, type B contracts may be purchased with or without a deductible, and type C contracts provide full coverage without a 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 their bonus-malus record

In this setting, opportunist policyholders may take advantage of manipulating claims for several reasons. Firstly, according to the premium recouping interpretation of Li et al. (2013), policyholders who wrongly pretend to have incurred some small losses in order to recoup part of their insurance premium are more likely to be among the policyholders who do not plan to keep a long term relationship with the same insurance company. Intuitively, such customers feel a lower moral cost of defrauding than those who intend to keep a long-term relationship with their insurer.[7] In our empirical analysis, this will lead us to define a *Recoup Group RG* that includes the policyholders who did not renew their contract more than one year after the policy year under consideration.[8] Secondly, for two reasons, insurance contracts may also incentivize opportunistic policyholders to manipulate claims corresponding to true accidents. Indeed, 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

[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 highlighted by Li et al. (2013) could reflect a kind of resentment against insurers, particularly from policyholders who have not filed a claim during the previous months the policy year.

[8]In Taiwan, filing a claim during the last month of the policy year does not affect the policyholder through the bonus-malus system if he/she does not stay more than one year with the same insurer. Our definition of the *Recoup Group* thus corresponds to policyholders without strong attachments to their current insurer, and for whom false claims filed toward the end of the policy year have no consequence through the bonus-malus system.

to renew their contract with the same insurer may see an advantage in postponing their claim to the last policy month, in order to delay the increase in premium.[9] In addition, since the bonus-malus record depends on the number of claims and not on their severity, policyholders may benefit from filing one single 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. In brief, because of the bonus-malus system and of deductible contracts, postponing the first claim and merging any other accident with the first one within a single claim is a winning strategy for opportunistic policyholders.[10]

Type A and B contracts are subject to such claim 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 other hand, 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 of year 2010, the set of policyholders who renewed type A or B contracts in 2011, but not in 2012, with the same insurer will be called the *Suspicious Group* SG because of this maximum incentive to manipulate the bonus-malus system, with subgroups $SG1$ and $SG2$ for no-deductible and deductible contracts, respectively.[11] In 2018, all type A or B contracts included a deductible, and thus the distinction between $SG1$ and $SG2$ is no longer appropriate for this year.[12]

One of the key insights of our analysis will be about the role of DOAs in this fraudulent claim manipulation process in 2010. Figure 2 provides a preliminary idea of this role by considering how the type of contract and the sale process (DOA or standard insurance

[9]In addition, the bonus-malus system forgives the first accident for drivers who have had no other accidents for three years, which provides an even larger manipulation gain.

[10]Since 1996, the per-claim deductible is increasing with the number of claims, which strengthen even more the incentives to manipulate claims.

[11]The bonus-malus record has a new departure point when the policyholder switch insurers. Thus, by postponing their first claim to the last policy month, such policyholders were able to fully escape the consequence of this claim on their bonus-malus record.

[12]In other words, SG is 2018 corresponds to $SG2$ in 2010.

agents) have affected the time distribution of claims during the policy year. It is striking how the claim distribution during the last policy month is peaking at the end of the year for members of $SG1$ and $SG2$ who have purchased insurance through agents sheltered by car dealers. Comparing with type C contracts used as a benchmark without claim manipulation reinforces the intuition that DOAs played an important role in this fraud process.

Figure 2

It nevertheless remains that Figure 2 does not allow us to assess whether this timing favored by DOAs resulted from the manipulation of claims corresponding to actual losses or from the behavior consisting in filing a small false claim at the end of the policy year in order to recoup some money from the insurer. However, if a substantial number of claims filed in the last policy month correspond in fact to first claims that have been postponed, possibly with the cumulated losses of two events, then such claims should be more costly than average. In other words, 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, contrary to the premium recouping interpretation of Li et al. (2013), hence a possible way of disentangling these two interpretations. But this may be misleading if the cost of claims is affected by an intertemporal moral hazard mechanism. Indeed, if a first accident makes drivers more cautious, then one may expect that subsequent accident would tend to be less severe, hence another possible explanation for an increase in the *first claim cost ratio* by the end of the policy year. To separate claim manipulation from moral hazard, we may consider type C contracts as a benchmark, since claim manipulation is vey unlikely for such contracts.

Figure 3

Figure 3 sustains the claim manipulation hypothesis for policyholders of the $SG2$ group who have purchased insurance through a DOA in 2010: their *first claim cost ratio* strongly increases in the last month of the policy year, and this is not the case for the other groups of policyholders. This suggests that in 2010 the claim manipulation mechanism dominated the premium recouping mechanism in $SG2$ (the subgroup of policyholders who benefit the most from claim manipulation), with DOAs acting as fraud facilitators, while the reverse occurs in the other subgroups. We also observe that for the RG group, the first claim cost ratio slightly increases when insurance has been purchased through the DOA channel, while it slightly decreases otherwise. This suggests that, among RG policyholders, the claim manipulation mechanism may be stronger than the premium recouping mechanism when insurance goes through DOAs. As we will see later, things have changed from 2010 to 2018.

3 Theoretical background

The model features the non-cooperative interaction between policyholders and insurers, in a costly state verification setting.[13] Consider a population of risk-averse drivers, whose type is defined by the couple (i, h) with $i \in \{D, A\}$ and $h \in \{1, 2\}$. Index i refers to the individuals' preference for a specific distribution channel through which they purchase insurance: DOA when $i = D$ or standard insurance agents when $i = A$.[14] Index h reflects the individual's degree of absolute risk aversion: $h = 1$ corresponds to a higher absolute risk aversion than $h = 2$. Assume that drivers may have either 0,1 or 2 accidents during

[13] See Picard (1996). For the sake of brevity, several aspects of the insurance market analysis are deliberately overlooked here. This particularly concerns 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.

[14] For the sake of simplicity and brevity, we do not analyze the reasons for which an individual may prefer to purchase insurance through a car dealer or through another distribution channel. Such preferences are likely to depend on many factors, such as the valuation of time saved by bundling the purchase of a new car and the taking out of an insurance policy, the repeated relationship between individuals and car dealers that also provide repair services, or the level of trust in car dealers.

the same policy year, with probability π_1 and π_2 for 1 and 2 accidents, respectively, and $\pi_1 + \pi_2 < 1$, and also that these probabilities are independent of 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 $d_{i2} > d_{i1} \geq 0$.[15]

Each accident may be severe or minor, and the corresponding claims small or large, with probability q_s or $q_m = 1 - q_s$, respectively, irrespective of the policyholder's type, and whether it is the first or second accident during the policy year. To simplify our analysis of fraud through claim manipulation, its is assumed that a large claim exactly doubles a small claim, with loss ℓ and 2ℓ, respectively. 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 that occured during the last policy month, should another minor accident occur later during the same policy year. Otherwise, the claim corresponding to the first minor accident will be denied because filed outside the permitted time. Fraud reduces the retained cost of the accidents 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 intends to stay with the same insurer at least during the next year. Fraud requires collusion with a car repairer, the policyholder and the repairer sharing the benefits according to their respective bargaining powers. If they are spotted defrauding, they have to pay a penalty (considered, for simplicity, as a fine to the government), and, in that case, the claim is fully denied.

Let us denote by α_{ih} and β_{ih} the fraud and audit mixed strategy of the policyholder and 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 small claim (when the corresponding

[15] For notational simplicity, we assume that the deductible is the same whether it is the first or second claim during the policy year.

minor accident occurs before the last policy month), with the intention to file a single large claim for two accidents during the last policy month, should another minor accident occur before the end of the year. Fraud is concentrated among those policyholders who are willing to stay with the same insurer at the end of the policy year because they are the ones who benefit the most through the bonus-malus mechanism.[16] β_{ih} is the probability that a large claim (filed by a type (i, h) policyholder) is audited by the insurer.[17] Such large claims correspond either to true severe accidents or to two minor accidents that have been fraudulently aggregated and postponed to the last month). We assume that audit allows the insurer to detect with certainty whether the claim has been manipulated or not.

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

$$C_{ih} = L - D_{ih} + FC_{ih} + AC_{ih}, \tag{1}$$

where L is the expected costs of accidents, D_{ih} is the cost retained by the policyholder (in the absence of claim manipulation), FC_{ih} is the cost of claim manipulation for the insurer and AC_{ih} is the audit cost.

L and D_{ih} are equal to the expected number of accidents per policyholder $\pi_1 + 2\pi_2$ multiplied by the weighted average loss per accident and by the deductible per accident,

[16] The policyholders who may benefit the most from defrauding through claim manipulation are those who have a first minor accident before the last month of their policy year and who do not intend to switch insurers. If these policyholders are just indifferent between defrauding and not-defrauding, as will be the case at the equilibrium of the policyholder-insurer interaction game presented in the following analysis, then the other policyholders will be detered from defrauding.

[17] 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 policy 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.

respectively. This gives

$$\begin{aligned} L &= (\pi_1 + 2\pi_2)[q_s \ell + 2q_m \ell] \\ &= (\pi_1 + 2\pi_2)(2 - q_s)\ell, \end{aligned}$$

and

$$D_{ih} = (\pi_1 + 2\pi_2)d_{ih}.$$

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 through the detection of falsified claims. DOAs have some bargaining power with insurers and they may intercede with the insurer when a claim is denied for fraud. This intervention is successful with some probability, and thus it decreases the financial benefit drawn by the insurer from spotting a defrauding policyholder-car repairer coalition. Thus, we may write

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

where $a_1(d_{ih})$ and $a_2(d_{ih}, \zeta_i)$ correspond to the expected cost of fraud (in the absence of audit), and to the expected gain from claim audit. 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. Furthermore, ζ_i is a parameter that measures the bargaining power of distribution channel i, with $\zeta_D > \zeta_A$.[18] We have $a_{2\zeta}' < 0$ because the distribution channel's bargaining power leads to a smaller insurer's

[18] Claim manipulation, as it is described, may be committed by policyholders who intend to renew their insurance policy and who have two accidents, the first one being minor and occurring before the last month of the policy year. Thus, $a_1(d_{ih})$ and $a_{2i}(d_{ih})$ depend on the probability that a type (i, h) individual is in this situation, which depends on π_1, π_2 and q_s, but also on the timing of accidents throughout the policy year, which is left undescribed for the sake of brevity.

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 manipulated 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 the policyholder-repairer collusion more difficult. We denote c_i the audit cost when the insurance distribution channel is $i = D$ or A, with $c_D > c_A$.

Since here fraud consists in filing one single large postponed claim for two accidents, the number of large claims filed by type (i, h) policyholders is linearly increasing with α_{ih}, which allows us to write[19]

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

The insurer chooses β_{ih} in $[0, 1]$ in order to minimize C_{ih} given by (1), which implies

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

where

$$\alpha^*(d, \zeta, c) \equiv \frac{c a_3}{a_2(d, \zeta) - c a_4}. \tag{5}$$

with $\alpha^{*\prime}_d > 0, \alpha^{*\prime}_\zeta > 0$ and $\alpha^{*\prime}_c > 0$. Let us assume that $\alpha^*(d, \zeta, c) < 1$ for the relevant values of d, ζ, c, which 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

[19] Here also, a_3 and a_4 depend on π_1, π_2 and q_s (but not on d_{ih}), and furthermore a_4 depends on the timing of accidents throughout the policy year.

(which is not explicitly defined here)[20], on the fine imposed on spotted defrauders, and on their degree of risk aversion, type h policyholders are willing to defraud if the probability of being caught is smaller than a threshold $\beta_h^*(P_{ih}, d_{ih}, \zeta_i) \in (0,1)$. Individuals always defraud when the audit probability is zero, and they never defraud if all large claims are audited: hence the audit probability $\beta_h^*(P_{ih}, d_{ih}, \zeta_i)$ for which they are indifferent between fraud and honesty is between 0 and 1.[21] 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,[22] and it is increasing with d_{ih} because an increase in the deductible makes fraud more attractive. Furthermore, β_h^* is increasing with ζ_i because a larger bargaining power of the agent corresponds to a larger probability of avoiding the full cancellation of the insurance payout when a fraudulent claim is detected through an audit. Thus, we have

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

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

[20] We may, for instance, assume that policyholders make take it or leave it offers to car repairers. The word "bribe" refers to any form of advantage that the car dealer-repairer firm may obtain from the arrangement with the policyholder, such as the guarantee of a future car purchase.

[21] β_h^* could be defined in a more explicit way by considering the expected utility of a type h individual who has a minor accident before the last policy month, and who has to choose between two strategies: either honestly filing a small claim without delay, or postponing her claim to the last policy month in order to file a single large claim if another minor accident occurs. β_h^* is the audit probability that makes the policyholder indifferent between these two strategies.

[22] For instance, under DARA preferences, 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 deterred.

$\alpha_{ih} = 1$ 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 $\beta_{ih} = \beta^*_h(P_{ih}, d_{ih}, \zeta_i) \in (0, 1)$ and $\alpha_{ih} = \alpha^*(d_{ih}, \zeta_i, c_i) \in (0, 1)$.

In brief, at equilibrium, the audit probability $\beta_{ih} = \beta^*_h(P_{ih}, d_{ih}, \zeta_i)$ makes the policyholder indifferent between manipulation and honesty, and the manipulation 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 claim manipulation. Firstly, using $\alpha^{*\prime}_d > 0$ shows that higher deductibles go along with more manipulation. Since $d_2 > d_1 \geq 0$, we have $\alpha_{i2} > \alpha_{i1}$ for $i \in \{D, A\}$. 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. Furthermore, using $c_D > c_A, \xi_D > \xi_A$, and $\alpha^{*\prime}_\zeta > 0, \alpha^{*\prime}_c > 0$ yields $\alpha_{Dh} > \alpha_{Ah}$ for $i \in \{1, 2\}$. Put briefly, for a given type of individual, there is more fraud when insurance has been purchased through the DOA agents than through standard insurance agents, either because it is more costly to audit a claim that goes through a DOA or because DOAs have a larger bargaining power than standard insurance agents.

4 Data and testing of hypotheses

4.1 The data

The data yielded by Companies 1 and 2 provides detailed information about the policyholders, their insurance contracts and the claims they have filed. Available variables are listed in Table 1. Data was collected over the 2010-2012 and 2018-2020 periods. However, our analysis of insurance claims will be restricted to 2010 and 2018, in order to know whether policyholders subsequently renewed their contracts for less or more than one year.[23] We will start by considering year 2010 in sections 4.2 and 4.3. As previously mentioned, Company 1 has strongly reduced its dependence on DOAs from 2010 to 2018, and thus in section 4.4 comparing results obtained for years 2010 and 2018 will allow us to appraise the consequence of this strutural change.

We target the owners of private usage small sedans and small trucks with type A, B or C insurance contracts for automobile physical damage. In 2010, there was $121,952$ policyholders in the sample, and 8.10% of them filed at least one claim, which corresponds to $9,874$ observations. This subset defines our "research sample", i.e. the sub-sample of policyholders with claims.

Tables 1 and 2

The mean values of the variables in the two samples are displayed in the first two columns of Table 2, with some significant differences. In particular, the percentages of type A or B contracts, and particularly those in the suspicious groups $SG1$ and $SG2$, are much larger in the research sample. The three other columns in Table 2 separate the research sample into three subgroups, according to the insurance distribution channels (DOA in Company 1 and non-DOA in Companies 1 and 2), with significant differences in terms of gender, usage, and vehicle size. There is also a much larger proportion of new vehicles for the DOA channel, which reflects the fact that, most of the time, a DOA sells an insurance contract when the corresponding dealer sells a new car. The percentage of claims filed during the last month of the policy year, measured by the average value of dummy SC, and the share of the RG group are larger in the DOA channel than in the two other channels.

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

4.2 Evidence on claim manipulation

Our first step consists in testing whether in 2010 the perspective of a one-year contract renewal and the choice of a deductible contract stimulate insurance fraud by postponing claims to the last policy month, called the "suspicious period", possibly by filing one claim for two events. In other words, we wonder whether belonging to the Suspicious Group SG, and particularly subgroup $SG2$, is a factor that has stimulated insurance fraud through claim manipulation. Defining the fraud rate as the number of claims per policyholder filed during the suspicious period[24] leads us to formulate the following hypothesis.

Hypothesis 1 (H1): *The fraud rate tends to be higher in the suspicious group than in the non-suspicious group, and this is particularly the case for individuals covered by deductible contracts.*

Testing **H1** amounts to identifying whether there is a conditional dependence between belonging to the suspicious group and filing a claim within the suspicious period, respectively associated with dummies SG (or $SG1$ and $SG2$ for each subgroup) and SC. We do so through the following three Bivariate Probit models, where $\Phi(.)$ is the cumulative normal distribution function, and X is the vector of explanatory variables (with vectors of coefficients $\beta_{SC}, \beta_{SG}, ...$), including the premium amount and all the variables used in pricing and underwriting decisions.[25] In order to control for the recouping effect, dummy RG is also included in X.

Model 1:

$$\text{Prob}(SC = 1) = \Phi(X\beta_{SC} + \varepsilon) \tag{7}$$

$$\text{Prob}(SG = 1) = \Phi(X\beta_{SG} + \eta) \tag{8}$$

[24] Of course, this definition of the fraud rate does not mean that all claims filed during the suspicious period have been fraudulently manipulated.

[25] This includes all the observable characteristics of the insured (e.g., age, gender, bonus-malus coefficient, premium, etc...), the characteristics of the vehicle (e.g., age, brand, registered area, etc...) and recoup dummy RG. Hence, X includes all the variables listed in the first part of Table 1, and *logprem* and RG in the second part.

Model 2:

$$\begin{aligned}
\text{Prob}(SC &= 1) = \Phi(X\beta_{SC} + \varepsilon) && (9)\\
\text{Prob}(SG1 &= 1) = \Phi(X\beta_{SG1} + \eta) && (10)
\end{aligned}$$

Model 3:

$$\begin{aligned}
\text{Prob}(SC &= 1) = \Phi(X\beta_{SC} + \varepsilon) && (11)\\
\text{Prob}(SG2 &= 1) = \Phi(X\beta_{SG2} + \eta) && (12)
\end{aligned}$$

The results of these regressions are presented in Table 3, with a special interest in the residual correlation coefficient ρ. **H1** should lead to a positive conditional correlation between filing a suspicious claim and belonging to a suspicious group. More formally, the estimated residual correlation coefficients of these models $\widehat{\rho}_{SC,SG}$, $\widehat{\rho}_{SC,SG1}$ and $\widehat{\rho}_{SC,SG2}$ should be positive and significantly different from 0, which leads us to test for the null hypothesis $H_0 : \rho_{SC,SG} \leq 0, H_0 : \rho_{SC,SG1} \leq 0$ and $H_0 : \rho_{SC,SG2} \leq 0$, in models 1, 2 and 3 respectively.

The three estimated residual correlation coefficients are significantly positive, which allows us to reject the null hypothis in each model, and thus to state that there is a significantly positive conditional correlation between SC and $SG, SG1$ or $SG2$, in each model. In other words, in accordance with **H1**, there is a conditional dependence between belonging to the suspicious group and filing a claim within the suspicious period, whether the individual is covered by a deductible contract or not.[26]

[26] Table 3 also offers some interesting byproducts that are worth mentioning. Firstly, the policyholders from the RG group tend to file their first claims in the suspicious period, which echoes the conclusions of Li et al. (2013). Secondly, 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.

Table 3

When manipulation consists in postponing claims to the suspicious period, by cumulating several losses in a single claim when possible (which differs from small claims filed by the end of the policy year to recoup a part of the insurance premium), then the suspicious period should be characterized by high values of the first-claim cost ratio. This is expressed in Hypothesis 2.

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

Hypothesis 2 is tested through the following regression:

$$clmamt = \alpha_c SC + \alpha_f first + \alpha_{fs} first * SC + \alpha_X X + e, \tag{13}$$

which is performed among the claims filed by members of $SG1$ and $SG2$ groups. This corresponds to 6,974 claims filed by 6,521 policyholders from SG1, and 695 claims filed by 647 policyholders from SG2. In these regressions, $clmamt$ is the value of the claim (in US dollars), while SC and $first$ are dummies indicating respectively that the claim was suspicious (i.e., it was filed during the last month of the policy year), and that it was the first claim of the policyholder during this policy year. Regression (13) also includes the interaction variable $first * SC$. Results are reported in Table 4.

Table 4

The estimated coefficients of the interaction variable are $\widehat{\alpha}_{fs} = -113.3$ with $p-$value 0.1627 for $SG1$, and $\widehat{\alpha}_{fs} = 1465.7$ with $p-$value lower than 0.0001 for $SG2$. This sustains Hypothesis 2 for $SG2$, but not for $SG1$, which confirms the fact that being covered by a deductible contract is a factor that stimulates fraud through claim manipulation. Hypothesis 3 focuses attention on the role of DOAs in this type of insurance fraud.

Hypothesis 3 (H3): *The fraud rate in the suspicious group is larger when insurance has been purchased through the DOA channel than through other distribution channels.*

We test **H3** by testing Bivariate Probit models 1, 2 and 3 in sub-samples that include the policyholders who purchased insurance through the DOA channel or through other distribution channels. This leads us to estimated residual correlation coefficients $\widehat{\rho}_{SC,SG}, \widehat{\rho}_{SC,SG1}$ and $\widehat{\rho}_{SC,SG2}$ in each subsample.

Tables 5,6 and 7

Detailed results are displayed in Tables 5, 6 and 7, for models 1, 2 and 3 respectively, with conclusions on residual correlation summarized as follows:[27]

		Company 1 Dealer	**Company 1 Non-dealer**	**Company 2**
Model 1	$\widehat{\rho}_{SC,SG}$	0.5393^{***}	0.1344	0.0562
Model 2	$\widehat{\rho}_{SC,SG1}$	0.5729^{***}	0.0916	-0.0610
Model 3	$\widehat{\rho}_{SC,SG2}$	0.7492^{***}	-0.2020	0.2076^{***}

Hence, when the regressions are performed in the sub-sample of policyholders who purchased coverage through the DOAs of Company 1, there is a significant positive residual correlation between SC and $SG, SG1$ or $SG2$ at the 1% threshold. This correlation vanishes in the two other sub-samples, except between SC and $SG2$ in Company 2.

4.3 Complements on the role of car dealers in claim manipulation

The previous conclusions may be reinforced by testing whether $\widehat{\rho}_{SC,SG}, \widehat{\rho}_{SC,SG1}$ and $\widehat{\rho}_{SC,SG2}$ are significantly larger among the policyholders who purchased insurance through car dealers than through other channels. To do so, we successively consider the two null

[27] *** refers to significance level at the 1% threshold.

hypotheses $H_0 : \widehat{\rho}^{D}_{SC,SG} \leq \widehat{\rho}^{ND}_{SC,SG}$ and $H_0 : \widehat{\rho}^{D}_{SC,SG} \leq \widehat{\rho}^{C2}_{SC,SG}$ in model 1, where D, ND and $C2$ refer respectively to insurance purchased from Company 1 through dealers, from Company 1 through other distribution channels, and from Company 2. We proceed in the same way for models 2 and 3, hence with $SG1$ and $SG2$ instead of SG. Results are displayed in Table 8. The two null hypotheses $\widehat{\rho}^{D}_{SC,SG} \leq \widehat{\rho}^{ND}_{SC,SG}$ and $\widehat{\rho}^{D}_{SC,SG} \leq \widehat{\rho}^{C2}_{SC,SG}$ are rejected at 1% significance level, and the conclusion is unchanged for $SG1$ and $SG2$. In other words, whatever the definition of the suspicious group ($SG, SG1$ or $SG2$), the conditional correlation between filing a suspicious claim and belonging to the suspicious group is significantly larger when contracts are sold through the car dealer associated with company 1 than through another distribution channel of company 1 or from company 2.

Table 8

Further evidence on the role of car dealers may be obtained by focusing attention on the first-claim cost ratio during the suspicious period (as in Hypothesis 2) by considering subsamples defined by the distribution channel, and by using type C contracts as a benchmark. A first-claim cost ratio during the suspicious period larger for $SG1$ or $SG2$ than for type C contracts would signal claim manipulation by members of the suspicious groups. Symmetrically, a lower first-claim cost ratio would be compatible with the premium recouping mechanism highlighted by Li et al. (2013), with small claims filed at the end of the policy year if no claim has been filed before. This leads us to consider regression (14) below, where the claim amount is the dependent variable as in regression (13). In (14), $first, SC$ and X are identical to those in regression (13), and S_1, S_2 and S_3 are dummies indicating that the policy has been purchased from Company 1 through the DOA channel, from Company 1 through another distribution channel and from Company 2, respectively. Furthermore C is a dummy indicating that the insurance policy is a type C contract, used as a benchmark without claim manipulation.

$$
\begin{aligned}
clmamt &= \alpha_c SC + \alpha_f first + \alpha_{fs} first * SC \\
&+ s_{SG11fs} SG1 * S_1 * first * SC \\
&+ s_{SG21fs} SG2 * S_1 * first * SC \\
&+ s_{SG23fs} SG2 * S_3 * first * SC \\
&+ s_{Cfs} C * first * SC + \alpha_X X + e. \qquad (14)
\end{aligned}
$$

The estimation of regression (14) shows that the null hypothesis $H_0 : s_{SG21fs} \leq s_{Cfs}$ is rejected at 1% significance level, contrary to the results obtained when s_{SG11fs} and s_{SG23fs} are compared to s_{Cfs}.[28] This means that the first-claim cost ratio is significantly higher during the last policy month when a deductible contract has been purchased from Company 1 through the DOA channel. All in all, in 2010 deductible contracts sold through DOAs have created the most favorable condition for insurance fraud through the postponing and aggregation of claims.

4.4 Smaller bargaining power for DOAs in 2018

From 2010 to 2018, Company 1 has cut almost by half the share of its automobile insurance contracts sold through car dealers. The latters became less important partners of the insurer, with presumably a lower bargaining power in the claim settlement process.[29]

To assess the consequences of this change, we have collected information about 269,475 automobile insurance contracts of type A, B and C, sold by Company 1 in 2018. The content of these contracts basically remained the same as in 2010, the only important

[28] Detailed estimation results are available from the authors upon request.

[29] In 2011, the largest car dealer group in Taiwan created its own insurance company, which induced other insurers to gradualy redirect a substantial part of their business toward other distribution channels. In the case of Company 1, the market share of DOAs decreased from 36.92% to 27.95%.

change being that in 2018 Company 1 only sold type A and B contracts with a deductible. Therefore, the Suspicious Group SG has no longer to be splitted between $SG1$ or $SG2$, and it coincides with what we called $SG2$ for year 2010. Table A1 in Appendix provides detailed information about the data. Comparing Tables 1 and A1 confirms the decrease in the proportion of contracts sold through DOAs, and other important changes including the decrease from 6.16% to 3.71% in the proportion of policyholders who filed a claim in 2010 and 2018, respectively. Figure A1 also confirms that claim rates are still higher during the last month than during the previous months of the policy year, with a large decrease in the average claim cost during the last policy month, and Figure A2 shows a decrease in the first claim cost ratio for all types of contracts, including those in SG going through DOAs, contrary to what was observed for $SG2$ in 2010.

Does this mean that the claim manipulation favored by DOAs has vanished in 2018? Formal tests have been performed to find out for sure. The results of Bivariate Probit regressions (similar to Model 1 above) are presented in Table A2. The estimated residual correlation between SG and SC is still significantly positive whatever the distribution channel, but the null hypothesis $H_0 : \rho^D \geq \rho^{ND}$ is rejected at the 1% significance threshold. In other words, the positive residual correlation between belonging to the suspicious group and filing a claim in the suspicious period still holds, which confirms claim manipulation, but the role of DOAs in this fraud process has vanished. For the sake of completeness, we have checked that the difference $\rho^D - \rho^{ND}$ has significantly decreased between 2010 and 2018, which means that the higher conditional correlation between SC and SG for the DOA channel, by comparison with other distribution channels, has significantly decreased from 2010 to 2018.[30]

We have performed a robustness check by a two-stage method in order to confirm this change from 2010 to 2018. To do so, we have created a new data set that includes $SG2$

[30]In more precise terms, the null hypothesis $H_0 : (\rho^D - \rho^{ND})_{2018} - (\rho^D - \rho^{ND})_{2010} \geq 0$ can be rejected at the 1% significance level.

and type C contracts sold by Company 1 in 2010 or 2018, with dummy y_{2018} used to indicate that the contract has been sold in 2018.[31] The first stage consists in estimating the following Probit regression:

$$\Pr[SG = 1] = \Phi(X\beta_{SG} + \eta),$$

and the estimated probability of belonging to the Suspicious Group $\widehat{SG}$ and dummy D for the DOA channel are used as explanatory variables in the second-stage regression:

$$\begin{aligned}\Pr[SC &= 1] = \Phi(\beta_{estSG}\widehat{SG} + \beta_{SG}SG + \beta_D D + \beta_{2018}y_{2018} \\ &+\beta_{SGD}SG * D + \beta_{SG2018}SG * y_{2018} + \beta_{D2018}D * y_{2018} \\ &+\beta_{SGD2018}SG * D * y_{2018} + X\beta_{SC} + \varepsilon)\end{aligned}$$

Results are presented in Table A3. The estimated coefficient of the triple interaction term $SG*D*y_{2018}$ is $\widehat{\beta}_{SGD2018} = -1.7265$, and it is significantly different from 0 at the 1% significance threshold. In other words, the stimulation effect of DOAs on the manipulation of claims by policyholders from the suspicious group SG has significantly decreased from 2010 to 2018.

Considering that DOAs played a crucial role in the manipulation of claims in 2010, one may wonder whether the decrease in their bargaining power has fully cancelled the fraud process in 2018, be it under the form of claim manipulation or of the premium recouping behavior. To get an idea, we have estimated regression (13) for the SG group and for the type C contracts, with the data of year 2018. Results are presented in Table A4. The estimated coefficient $\widehat{\alpha}_{fs}$ is not significantly different from 0 in the two subsets of contracts. In other words, in 2018, contrary to what occured in 2010, there was no significant change in the average amount of the claims filed during the last month of the

[31] $SG2$ and SG coincide in 2018 since Company 1 has stopped selling type A or B contracts without deductible.

policy year by comparison with previous months.

5 Conclusion

The purpose of this paper was to analyze some aspects of the policyholder-service provider coalition in insurance fraud mechanisms: how it can affect the credibility of claim auditing, and how fraudulent claim manipulation may emerge. It is a fact that the economic analysis of insurance fraud is often based on a very abstract picture of claim 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 car insurance case offers such a possibility, with fraud also 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. The policyholders with deductible contract who intend to renew their policies (the suspicious group) have a larger propensity to defraud in that way than other policyholders

Our main focus was on the role of DOAs in this fraud process, with two specificities for this distribution channel. Firstly, the collusion between car repairers and policyholders is easier when insurance agents and car repairers are sheltered by a car dealer, and establishing claim manipulation unambiguously is more costly (i.e. the audit cost is larger) in that case. Secondly, DOAs may more easily escape penalties when fraud is detected (i.e. their bargaining power is larger at the claim settlement stage) because they can retaliate by redirecting their customers toward other insurers if the relationship with the current insurer deteriorates. Both specificities are related to the multi-faceted activities of DOAs: they sell insurance contracts, but they also work hand in hand with car repairers and car dealers. The comparison between years 2010 and 2018 suggests that reducing the dependence on car dealers has allowed Company 1 to deter claim manipulation more efficiently, because of the decrease in the bargaining power of DOAs. In other words, the role of DOAs in car insurance fraud seems to be much more related to their bargaining power, than to the difficulty for the insurer to establish that claims had been manipulated.

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.

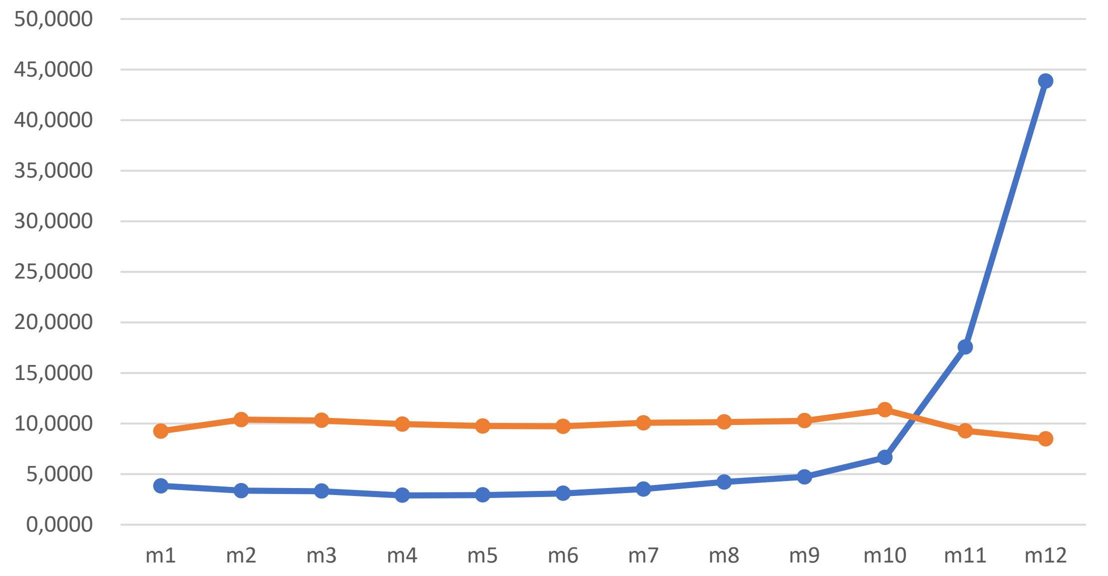
Pao, T.I., L.Y. Tzeng and K.C. Wang (2014), "Typhoons and opportunistic fraud: claim patterns of automobile theft insurance in Taiwan", *Journal of Risk and Insurance*, 81(1): 91-112.

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.

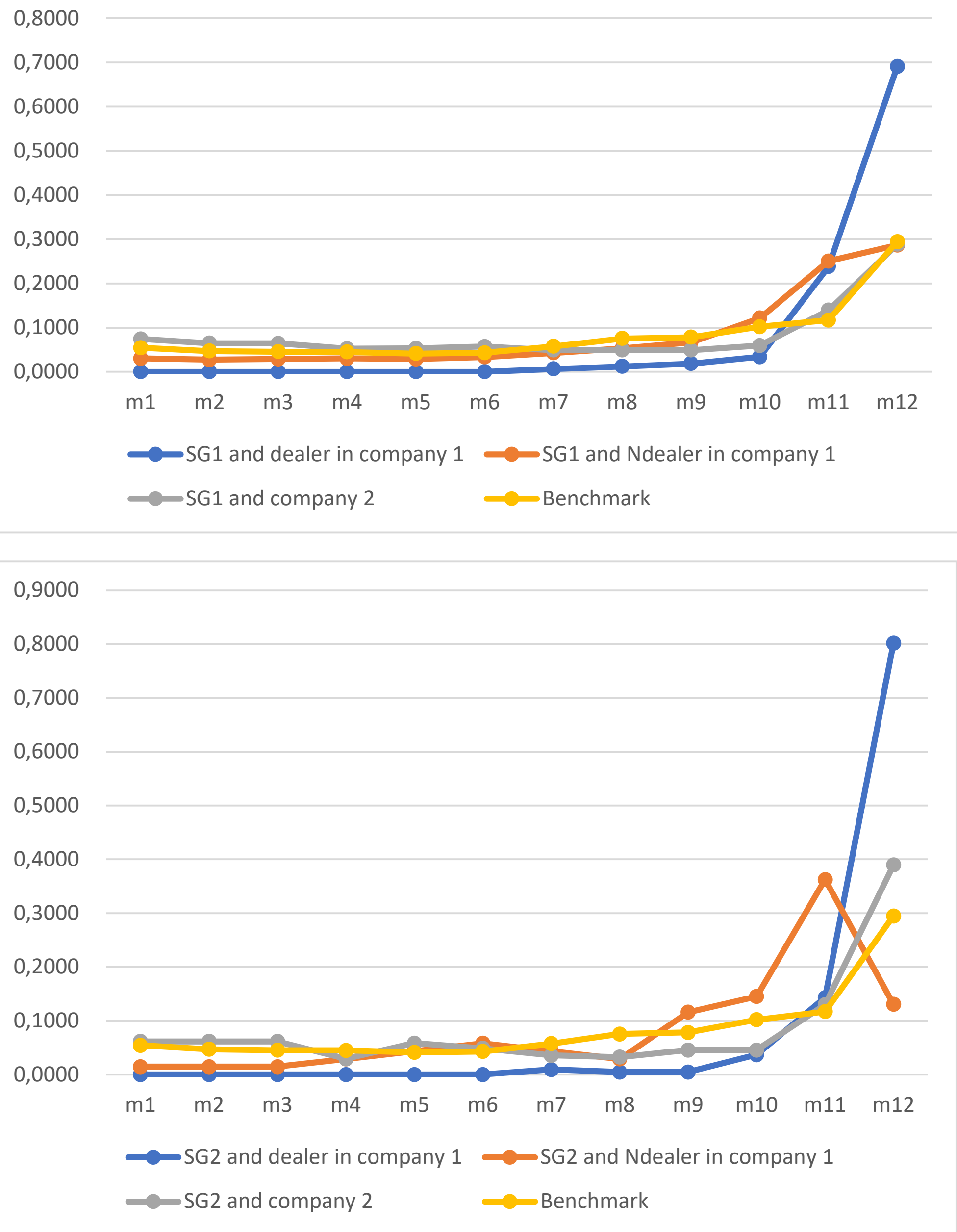
Figure 1: Distribution of claims and average cost of first claims during the policy year (2010)



Distribution of claims (unit : %) :

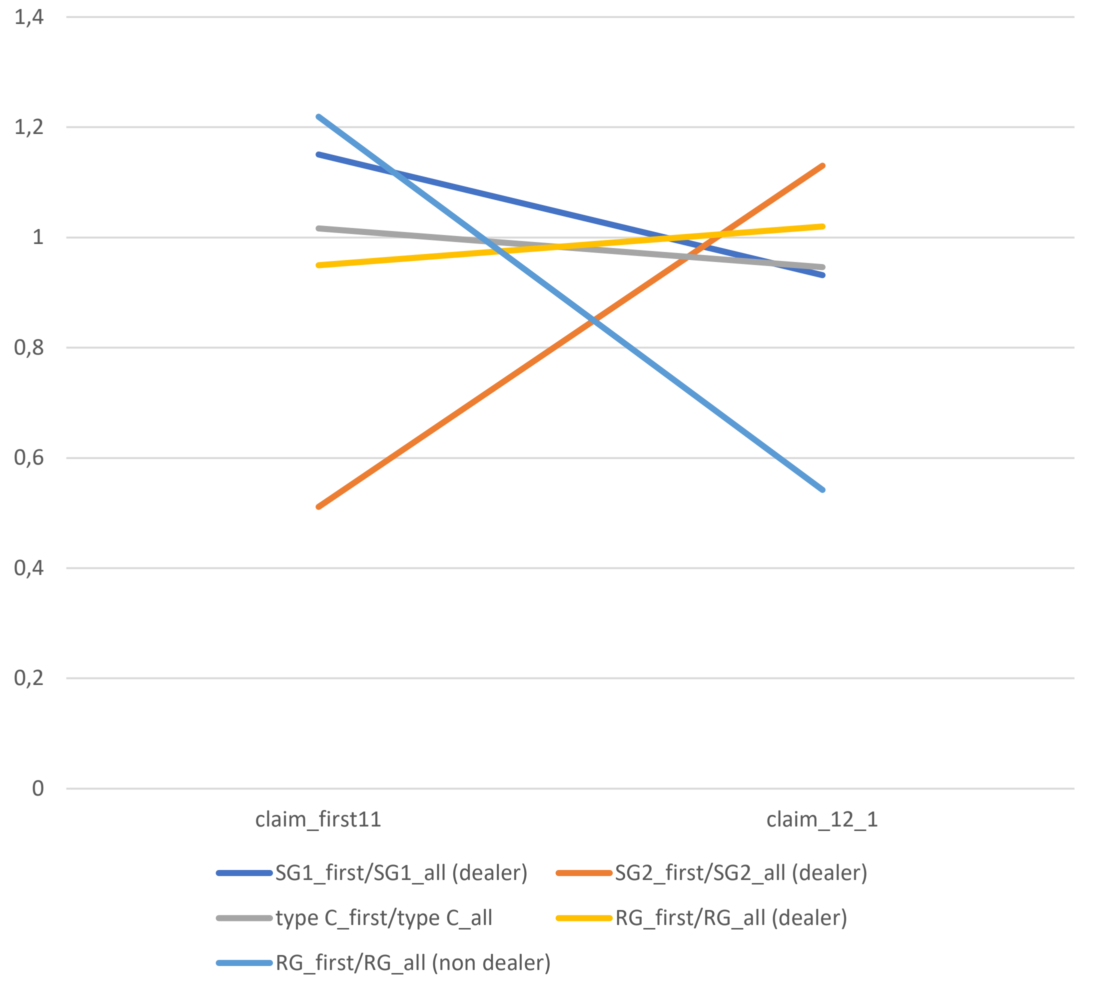
Average cost of first claims (unit : USD100) :

Figure 2: Distribution of claims during the policy year (2010) for *SG1*,*SG2* according to the sale process, with type C contracts as benchmark.



Qualifiers « dealer » and « Ndealer » refer to the cases where the insurance policy has been sold through a DOA or through a standard agent, respectively.

Figure 3: Changes in the First Claim Cost Ratio from months 1-11 to month 12 of the policy year (2010)



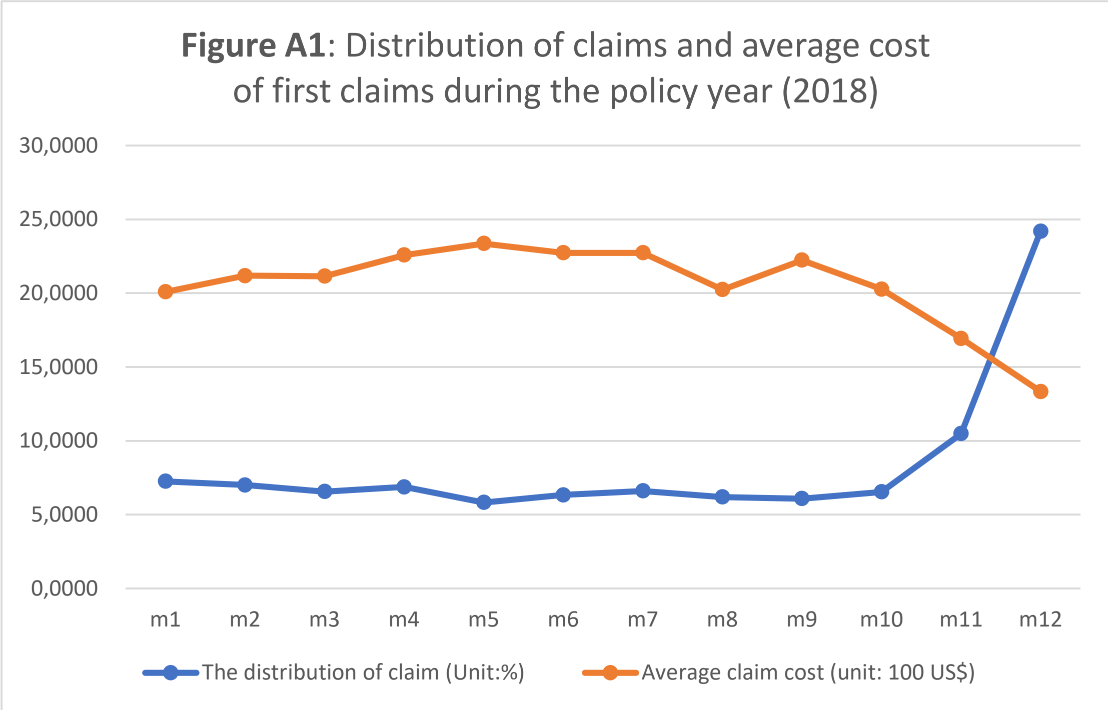


Figure A1: Distribution of claims and average cost of first claims during the policy year (2018)
30,0000
25,0000
20,0000
15,0000
10,0000
5,0000
0,0000
m1
m2
m3
m4
m5
m6
m7
m8
m9
m10
m11
m12
The distribution of claim (Unit:%)
Average claim cost (unit: 100 US$)

Figure A2: Changes in the First Claim Cost Ratio from months 1-11 to month 12 of the policy year (2018)

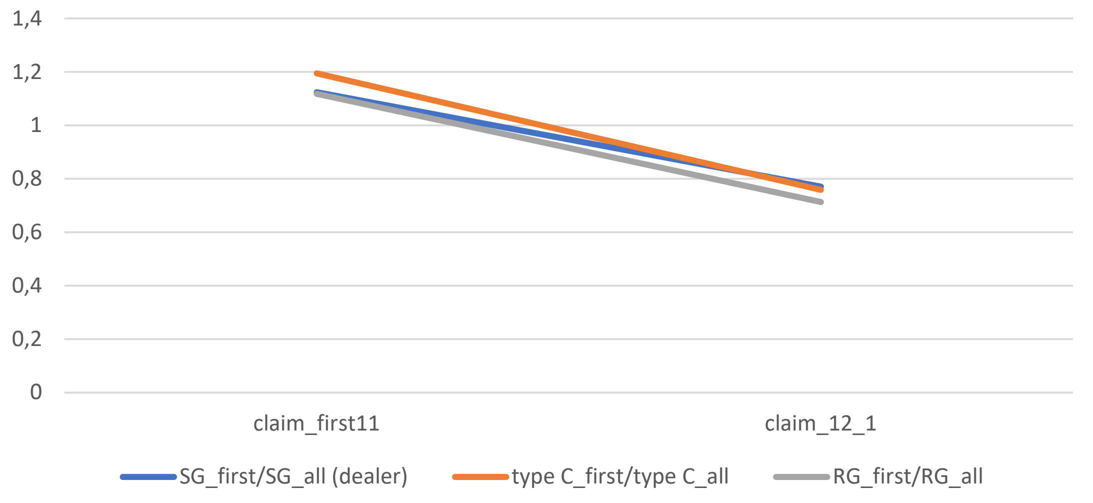


Table 1: Definition of variables

Variable	Definition
Explained variables:	
claim	Dummy variable equal to 1 when the insured has filed at least one claim during the policy year, 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.
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.
Explanatory variables:	
first group (underwriting and pricing variables)	
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 five 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.
sedan	Dummy variable equal to 1 when the car is a sedan and is for non-commercial

[1] The "suspicious group" (*SG*) includes the individuals who renew their contract with the same insurance company for only one year. The counter group for SG includes the policyholders who do not renew their contract, or renew their contract for more than one year with the same insurance company.
[2] The "suspicious group 1" (*SG*1) are the *SG*-group policyholders who also purchased the no-deductible contracts. The counter group for *SG1* includes the policyholders with deductible contract or who are not in *SG*-group.
[3] The "suspicious group 2" (*SG*2) are the *SG*-group policyholders who also purchased the deductible contracts. The counter group for *SG2* includes the policyholders with no-deductible contract or who are not in *SG*-group.

	or for long-term rental purposes, and 0 otherwise.[4]
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.
tramak_j	Dummy variable equal to 1 when the brand of the insured car is *j*, with *j*=*n*, *f*, *h*, *t*, *c*, and 0 otherwise.[5]
second group (other control variables)	
logprem	Logarithm of the premium of the contract in the current contract year.
D	Dummy variable equal to 1 if the insurance contract is sold through the DOA channel of company 1, and 0 otherwise.
company2	Dummy variable equal to 1 if the insurance contract is sold by company 2, and 0 otherwise. [6]
AB	Dummy variable equal to 1if the insured is covered by a type_A or type_B contract, and 0 otherwise.[7]
RG	Dummy variable equal to 1 when the insured belongs to the "recoup group",[8] and 0 otherwise.

[4] The counter group includes the insured cars are not small sedan, for example small or large truck, cargo…etc.

[5] The insured cars in counter group for *tramak_j*, *j*= *n*, *f*, *h*, *t*, *c* , are other brands (other than Nissan, Ford, Honda, Toyota, and China.)

[6] The contracts in counter group for *D* and *company 2* are the insurance contarcts sold through the channels other than DOA of *company 1*.

[7] The contracts in counter group for *type_A* and *type_B* are type_C contracts.

[8] The "recoup group" includes the policyholders who are in "recoup group" include the ones who are covered by type A or type B contracts and who do not renew their contract or renew it for only one year.

Table 2: Sample structure (2010)

	Whole sample	Sub-sample with claim	DOA in Company 1	Non-DOA in company 1	Company 2
claim	0.8100				
SC	0.0616	0.4386	0.6628	0.2723	0.2954
RG	0.2285	0.2365	0.3165	0.2197	0.1739
AB	0.3719	0.7316	0.8979	0.6175	0.6228
C2	0.2751	0.4741	0.0000	0.0000	1.0000
SG	0.1982	0.7316	0.8979	0.6175	0.6228
SG1	0.1748	0.6670	0.8386	0.5589	0.5522
SG2	0.0234	0.0645	0.0593	0.0585	0.0706
D	0.3692	0.3978	1.0000	0.0000	0.0000
female	0.7149	0.7436	0.7758	0.7176	0.7236
age2025	0.0024	0.0022	0.0022	0.0025	0.0021
age2530	0.0313	0.0386	0.0317	0.0339	0.0456
age3060	0.8930	0.8943	0.8965	0.8872	0.8944
ageabv60	0.0734	0.0650	0.0696	0.0763	0.0580
carage0	0.1947	0.2983	0.4926	0.1383	0.1785
carage1	0.1562	0.2403	0.2387	0.2214	0.2468
carage2	0.0951	0.1010	0.0688	0.0882	0.1315
carage3	0.1175	0.1117	0.0699	0.1272	0.1425
carage4	0.1041	0.0749	0.0440	0.0941	0.0956
veh_m	0.2912	0.2589	0.2283	0.2807	0.2786
veh_l	0.2580	0.2678	0.2701	0.3070	0.2553
sedan	0.9102	0.9247	0.9612	0.8974	0.9015
lnprem	9.0442	9.5277	10.0894	9.5279	9.0563
bonus	0.8954	1.1140	0.8760	0.7154	1.4214
Observations	149452	9205	3662	1179	4364

Table 3: Conditional dependence between SC and SG

Variables	Model 1		Model 2		Model 3	
	SC	***SG***	***SC***	***SG1***	***SC***	***SG2***
RG	0.1193***	1.5444***	0.1374***	1.3111***	0.3037***	2.1846***
	[0.0332]	[0.0861]	[0.0348]	[0.0732]	[0.1076]	[0.1744]
female	-0.0138	0.1493***	0.0034	0.3138***	-0.0299	0.0263
	[0.0317]	[0.0386]	[0.0330]	[0.0393]	[0.0524]	[0.0687]
age2025	0.0281	-0.1512	0.1413	-0.1264	0.2296	-0.8458
	[0.2993]	[0.3506]	[0.3045]	[0.3507]	[0.4003]	[0.6399]
age2530	-0.2138**	-0.2505**	-0.4537***	-0.2802***	-0.2659*	-0.7943***
	[0.0884]	[0.1071]	[0.0926]	[0.1083]	[0.1464]	[0.2049]
age3060	0.0409	0.0152	-0.0292	0.0367	-0.0627	-0.2297*
	[0.0551]	[0.0680]	[0.0568]	[0.0685]	[0.0964]	[0.1213]
tramak_n	0.1442	0.3462	0.1815	0.4390*	0.5884*	0.3559
	[0.1662]	[0.2265]	[0.1774]	[0.2277]	[0.3364]	[0.3788]
tramak_f	-0.1785***	0.0285	-0.1999***	0.0592	-0.0827	0.0165
	[0.0623]	[0.0749]	[0.0656]	[0.0769]	[0.0979]	[0.1249]
tramak_h	-0.1205**	-0.2087***	-0.0468	-0.1026	-0.1615*	-0.3503***
	[0.0566]	[0.0652]	[0.0582]	[0.0664]	[0.0897]	[0.1265]
tramak_t	0.0409	0.2473***	0.0697**	0.3562***	-0.0694	-0.2450***
	[0.0317]	[0.0392]	[0.0334]	[0.0400]	[0.0563]	[0.0748]
tramak_c	-0.4594***	-0.1594*	-0.3729***	-0.2453***	-0.2362**	-0.6231***
	[0.0765]	[0.0818]	[0.0784]	[0.0829]	[0.1086]	[0.1738]
carage0	0.3822***	0.4696***	0.3307***	0.4260***	0.4480***	0.5117***
	[0.0506]	[0.0600]	[0.0536]	[0.0611]	[0.870]	[0.1023]
carage1	0.1381***	0.0837	0.1361***	0.1840***	0.0998	0.3499***
	[0.0469]	[0.0532]	[0.0492]	[0.0547]	[0.0758]	[0.0949]
carage2	0.0573	-0.0801	0.0060	0.0526	0.1059	0.1412
	[0.0553]	[0.0626]	[0.0576]	[0.0643]	[0.0865]	[0.1164]
carage3	0.0956*	0.0376	-0.0547	0.0272	0.1182	-0.1662
	[0.0529]	[0.0595]	[0.0551]	[0.0609]	[0.0793]	[0.1176]
carage4	-0.1447**	-0.1928***	-0.2558***	-0.1329*	-0.2502***	0.0632
	[0.0607]	[0.0659]	[0.0637]	[0.0681]	[0.0901]	[0.1190]
veh_m	0.0783**	-0.2005***	0.1401***	-0.2903***	0.1156*	0.1263
	[0.3456]	[0.0421]	[0.0357]	[0.0430]	[0.0596]	[0.0806]
veh_l	0.0636	-0.0568	0.0544	-0.1804***	0.1670**	0.3927***
	[0.0403]	[0.0507]	[0.0418]	[0.0519]	[0.0771]	[0.0922]

sedan	0.0685	-0.3449***	0.0246	-0.2958***	-0.0286	0.0040
	[0.0585]	[0.0695]	[0.0605]	[0.0700]	0.0951]	[0.1277]
lnprem	0.1036***	0.4852***	0.1086***	0.4849***	0.0006	0.3405***
	[0.0257]	[0.0238]	[0.0272]	[0.0245]	[0.0426]	[0.0436]
bonus	-0.4794***	-0.1689***	-0.5341***	-0.1805***	-0.1382**	0.0550
	[0.0345]	[0.0400]	[0.0371]	[0.0415]	[0.0559]	[0.0682]
ρ	0.1395***		0.0873***		0.2608***	
	[0.0319]		[0.0337]		[0.0514]	

Standard errors in brackets; ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

Table 4: Testing hypothesis 2 (year 2010)

	SG1		***SG2***	
	Est. coeff	***P value***	***Est. coeff***	***P value***
Intercept	-2869.3	<.0001	-4642.1	<.0001
SC	-198.9	0.0113	-742.1	0.0183
first	46.5	0.4172	-403.0	0.0630
*first*SC*	-113.3	0.1627	1465.7	<.0001
female	17.1	0.4871	-145.3	0.0853
age2025	-237.2	0.3869	621.7	0.5280
age2530	-107.3	0.1077	-442.5	0.0865
age3060	-36.9	0.3693	223.1	0.1660
tramak_n	-201.7	0.0932	-620.8	0.1240
tramak_f	-184.5	0.0002	-134.0	0.3972
tramak_h	-117.8	0.0082	-138.6	0.4594
tramak_t	-193.9	<.0001	-401.3	<.0001
tramak_c	-219.1	0.0003	-836.5	0.0006
carage0	-149.2	0.0002	-108.0	0.3834
carage1	-103.0	0.0069	-192.3	0.1308
carage2	-25.8	0.5639	-159.7	0.2931
carage3	12.9	0.7677	-192.7	0.2024
carage4	103.0	0.0409	-30.5	0.8493
veh_m	-14.2	0.5944	-151.1	0.1689
veh_l	214.9	<.0001	148.3	0.1818
sedan	269.6	<.0001	305.2	0.0785
logprem	371.2	<.0001	697.1	<.0001
bonus	48.1	0.0681	-536.6	<.0001
Adj. R^2	0.1138		0.4206	
observations	6567		633	

Table 5: Conditional dependence between *SC* and *SG* in sub-samples – Model 1 (year 2010)

	Company 1 dealer		***Company 1 non-dealer***		***Company 2***	
	SC	***SG***	***SC***	***SG***	***SC***	***SG***
RG	0.2087***	1.4486***	-0.0127	1.2273***	0.1663***	1.0515***
	[0.0490]	[0.1864]	[0.1068]	[0.1624]	[0.0560]	[0.0861]
female	0.0602	0.1776**	-0.1042	0.1454	-0.0456	0.2719***
	[0.0535]	[0.0790]	[0.0914]	[0.1007]	[0.0463]	[0.0507]
age2025	0.3982	-0.3157	-0.2333	-0.4811	0.0419	-0.3576
	[5405]	[0.6116]	[0.9014]	[0.8294]	[0.4649]	[0.4989]
age2530	-0.4620***	-0.2670	0.0615	-0.6889**	-0.3168**	-0.5582***
	[0.1456]	[0.2155]	[0.2624]	[0.3040]	[0.1310]	[0.1383]
age3060	-0.0284	0.0917	0.0773	-0.2390	-0.0226	0.0069
	[0.0856]	[0.1254]	[0.1563]	[0.1686]	[0.0869]	[0.0947]
tramak_n	-0.0162	0.0922	0.5034	0.4999	0.3149	0.2470
	[0.3307]	[0.4795]	[0.4041]	[0.5685]	[0.2239]	[0.2816]
tramak_f	-0.0451	0.1857	-0.0466	-0.0111	-0.0337	-0.1600
	[0.1189]	[0.1755]	[0.1517]	[0.1682]	[0.0906]	[0.0993]
tramak_h	-0.0381	0.0760	-0.0707	-0.0458	-0.0103	-0.3598***
	[0.1253]	[0.1689]	[0.1447]	[0.1634]	[0.0750]	[0.0805]
tramak_t	-0.1260**	0.4904***	-0.1435	0.2018*	0.0418	-0.0453
	[0.0538]	[0.0760]	[0.0935]	[0.1037]	[0.0491]	[0.0539]
tramak_c	-0.1936	0.3031	0.1873	-0.3445	-0.1218	-0.3294***
	[0.3462]	[0.4113]	[0.2121]	[0.2479]	[0.0872]	[0.0937]
carage0	-0.0363	0.4796***	-0.1329	0.6156***	0.0914	0.4867***
	[0.0975]	[0.1268]	[0.1535]	[0.1729]	[0.0809]	[0.0934]
carage1	0.0224	0.1614	-0.1268	0.4121***	0.0151	0.2196***
	[0.0943]	[0.1219]	[0.1240]	[0.1320]	[0.0695]	[0.0703]
carage2	-0.1635	-0.1748	0.2890*	0.4076**	-0.0292	0.1062
	[0.1126]	[0.1472]	[0.1509]	[0.1756]	[0.0754]	[0.0790]
carage3	0.0092	-0.0060	-0.0056	0.1812	-0.1602**	0.1101
	[0.1111]	[0.1433]	[0.1332]	[0.1457]	[0.0719]	[0.0757]
carage4	-0.2389*	-0.2915*	-0.3620**	0.2705*	-0.1641**	0.0986
	[0.1254]	[0.1540]	[0.1597]	[0.1561]	[0.0810]	[0.0848]
veh_m	-0.0862	-0.2110**	-0.1666*	-0.1394	0.1415**	0.0495
	[0.0585]	[0.0836]	[0.1006]	[0.1110]	[0.0564]	[0.0618]

veh_l	-0.1789***	-0.2330**	-0.3471***	-0.0786	0.0888	0.2930***
	[0.0689]	[0.0926]	[0.1242]	[0.1263]	[0.0828]	[0.0909]
sedan	-0.1914	-0.2939	0.0666	-0.3395*	0.0700	-0.1867**
	[0.1258]	[0.1809]	[0.1567]	[0.1755]	[0.0816]	[0.0884]
lnprem	0.2229**	0.6791***	-0.0554	0.7074***	-0.0662	0.5853***
	[0.0874]	[0.0632]	[0.1058]	[0.0752]	[0.0453]	[0.5853]
bonus	-0.2188	-1.2374***	0.4098*	-1.0545***	-0.1504*	-0.4616***
	[0.1636]	[0.1890]	[0.2408]	[0.2304]	[0.080]	[0.0955]
Constant	-4.9188*	-4.9188***	-0.2026	-5.6378***	0.1742	-4.6010***
	[0.7791]	[0.5876]	[0.9081]	[0.6592]	[0.3822]	[0.3136]
ρ	0.5393***		0.1344		0.0562	
	[0.0729]		[0.1201]		[0.0480]	

Standard errors in brackets; ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

Table 6: Conditional dependence between *SC* and *SG1* in sub-samples – Model 2 (year 2010)

	Company 1 dealer		***Company 1 non-dealer***		***Company 2***	
	SC	***SG1***	***SC***	***SG1***	***SC***	***SG1***
RG	0.2010***	1.5341***	-0.0221	1.1578***	0.0709	1.2021***
	[0.0501]	[0.2253]	[0.1104]	[0.1645]	[0.0599]	[0.0980]
female	0.1167**	0.2759***	-0.0620	0.1638	-0.1032**	0.2954***
	[0.0549]	[0.0824]	[0.0935]	[0.1028]	[0.0480]	[0.0537]
age2025	0.4237	-0.4132	-0.1955	-0.1888	-0.3207	-0.2477
	[0.5564]	[0.6221]	[0.8616]	[0.8311]	[0.4869]	[0.5016]
age2530	-0.4478***	-0.5637**	-0.0376	-0.4613	-0.2080	-0.4562***
	[0.1503]	[0.2213]	[0.2680]	[0.3076]	[0.1329]	[0.1452]
age3060	-0.0198	0.0052	0.0910	-0.2645	-0.0325	-0.1003
	[0.0884]	[0.1334]	[0.1599]	[0.1707]	[0.0892]	[0.0987]
tramak_n	0.3639	0.0218	0.6697*	0.3062	0.3782	0.1518
	[0.3716]	[0.4651]	[0.4006]	[0.5392]	[0.2385]	[0.2967]
tramak_f	-0.1312	0.1808	0.0723	-0.1424	-0.2251**	-0.2843***
	[0.1246]	[0.1894]	[0.1563]	[0.1727]	[0.0958]	[0.1054]
tramak_h	-0.0022	0.1606	0.0368	-0.2072	-0.0244	-0.4242***
	[0.1272]	[0.1764]	[0.1485]	[0.1682]	[0.0769]	[0.0844]
tramak_t	-0.1027*	0.5401***	-0.0549	-0.0003	0.0102	-0.0394
	[0.0572]	[0.0802]	[0.0972]	[0.1056]	[0.0505]	[0.0565]
tramak_c	-0.3570	0.1435	0.0864	-0.4660*	-0.1067	-0.4536***
	[0.3465]	[0.4135]	[0.2209]	[0.2552]	[0.0892]	[0.0979]
carage0	-0.0659	0.5069***	-0.0428	0.5662***	0.0966	0.6870***
	[0.1021]	[0.1329]	[0.1574]	[0.1766]	[0.0888]	[0.0882]
carage1	0.0044	0.2771**	-0.0676	0.4466***	0.1138	0.4128***
	[0.0987]	[0.1267]	[0.1287]	[0.1354]	[0.0749]	[0.0739]
carage2	-0.0842	0.0655	0.2540	0.4376**	0.0698	0.2247***
	[0.1171]	[0.1550]	[0.1568]	[0.1804]	[0.0793]	[0.0830]
carage3	-0.0432	0.0676	0.0953	0.2778*	0.0236	0.2375***
	[0.1142]	[0.1472]	[0.1352]	[0.1472]	[0.0745]	[0.0796]
carage4	-0.3611***	-0.1421	-0.4584***	0.3412**	-0.0691	0.1746*
	[0.1289]	[0.1602]	[0.1689]	[0.1630]	[0.0842]	[0.0896]
veh_m	0.0165	-0.2147**	-0.1595	-0.3025***	0.1506***	0.0094
	[0.0596]	[0.0864]	[0.1011]	[0.1143]	[0.0584]	[0.0652]

veh_l	-0.1736**	-0.2738***	-0.3776	-0.2738**	0.1270	0.2227**
	[0.0699]	[0.0971]	[0.1246]	[0.1308]	[0.0866]	[0.0968]
sedan	-0.1373	-0.3944**	0.0408	-0.4320**	0.0674	-0.3254***
	[0.1279]	[0.1921]	[0.1612]	[0.1846]	[0.0838]	[0.0924]
lnprem	0.2181**	0.7433***	-0.0377	0.7134***	-0.0475	0.6283***
	[0.0890]	[0.0660]	[0.1169]	[0.0781]	[0.0484]	[0.0385]
bonus	-0.1648	-1.0223***	0.4459	-1.0166***	-0.2510***	-0.4993***
	[0.1718]	[0.1953]	[0.2571]	[0.2367]	[0.0932]	[0.1022]
Constant	-1.5136*	-5.7331***	-0.4028	-5.4420***	0.1944	-4.8641***
	[0.7873]	[0.6162]	[1.0080]	[0.6751]	[0.4086]	[0.3291]
ρ	0.5729***		0.0916		-0.0610	
	[0.0699]		[0.1362]		[0.0534]	

Standard errors in brackets; ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

Table 7: Conditional dependence between *SC* and *SG2* in sub-samples – Model 3 (year 2010)

	Company 1 dealer		***Company 1 non-dealer***		***Company 2***	
	SC	***SG2***	***SC***	***SG2***	***SC***	***SG2***
RG	0.2622	1.6147***	-0.4441	2.3649***	0.2728*	1.7827***
	[0.2004]	[0.2621]	[0.4225]	[0.4639]	[0.1594]	[0.2152]
female	-0.1535	-0.0666	-0.0555	0.3131	-0.0175	-0.0709
	[0.1249]	[0.1494]	[0.1397]	[0.2001]	[0.0669]	[0.0833]
age2025	0.2937	-0.4864	0.2417	0.4348	-0.0347	-0.2253
	[0.7007]	[0.8008]	[1.1300]	[1.0916]	[0.6076]	[0.7808]
age2530	-0.6705	-0.3919	0.6021	-0.7861	-0.3573*	-0.4311*
	[0.3223]	[0.3941]	[0.4272]	[0.7589]	[0.1828]	[0.2376]
age3060	0.0564	-0.1246	0.4362*	-0.1269	-0.1732	-0.1471
	[0.2093]	[2562]	[0.2572]	[0.3146]	[0.1268]	[0.1547]
tramak_n	0.7126	-0.1096	-0.3747	0.0300	0.2852	0.2575
	[0.6845]	[0.7266]	[1.7463]	[1.7441]	[0.4113]	[0.4415]
tramak_f	-0.1230	0.3446	-0.0555	0.0392	-0.0269	-0.0740
	[0.2342]	[0.2781]	[0.2220]	[0.2885]	[0.1283]	[0.1624]
tramak_h	-0.0649	-0.1577	0.0791	-0.5671*	0.0696	-0.2862**
	[0.2789]	[0.3410]	[0.2231]	[0.3424]	[0.1063]	[0.1434]
tramak_t	-0.4971***	-0.8100***	-0.1005	-0.1753	0.0083	-0.1858**
	[0.1352]	[0.1601]	[0.1473]	[0.1983]	[0.0727]	[0.0946]
tramak_c	0.0140	-0.0334	-0.2963	-0.2935	-0.1097	-0.8299***
	[0.5654]	[0.8117]	[0.2935]	[0.4233]	[0.1245]	[0.1970]
carage0	0.1719	0.4071*	-0.0541	0.2479	0.3881***	0.3900***
	[0.1898]	[0.2236]	[0.2660]	[0.3513]	[0.1199]	[0.1319]
carage1	-0.0991	0.0391	-0.0870	0.4110	0.0899	-0.0666
	[0.1911]	[0.2321]	[0.1933]	[0.2520]	[0.0954]	[0.1193]
carage2	-0.0875	-0.0744	0.2821	0.3666	0.0783	-0.0445
	[0.2321]	[0.2908]	[0.2407]	[0.3243]	[0.1058]	[0.1366]
carage3	0.3269	-0.1232	0.0080	-0.2482	0.2198**	-0.0513
	[0.2303]	[0.2926]	[0.2028]	[0.3401]	[0.0963]	[0.1284]
carage4	-0.2220	-0.4267	-0.3660	0.4180	0.0271	-0.0476
	[0.2400]	[0.3248]	[0.2303]	[0.2698]	[0.1083]	[0.1430]
veh_m	0.0385	0.0276	0.1005	-0.1990	0.1787**	-0.0262

	[0.1383]	[0.1680]	[0.1489]	[0.2141]	[0.0815]	[0.1075]
veh_l	0.1327	0.2949	0.1798	-0.0297	0.2780**	0.3828***
	[0.1801]	[0.1809]	[0.1944]	[0.2319]	[0.1228]	[0.1455]
sedan	-0.2460	-0.2723	0.0885	-0.6997**	0.1115	-0.3116**
	[0.2635]	[0.3166]	[0.2384]	[0.2843]	[0.1199]	[0.1471]
lnprem	0.1987	0.6067***	-0.1528	0.4173***	-0.0067	0.3557***
	[0.1692]	[0.1097]	[0.2006]	[0.1440]	[0.0713]	[0.0688]
bonus	-0.0319	-0.9141**	0.2821	-0.9554**	-0.3427**	0.0301
	[0.3338]	[0.3744]	[0.4282]	[0.4655]	[0.1385]	[0.1585]
Constant	-1.4106	-5.0827***	0.0842	-3.8736***	-0.1057	-3.6732***
	[1.4319]	[1.0026]	[1.6199]	[1.2197]	[0.5649]	[0.5400]
ρ	0.7492***		-0.2020		0.2076***	
	[0.1355]		[0.2206]		[0.0702]	

Standard errors in brackets; ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

Table 8: Difference of conditional dependence between *SC* and *SG/SG1/SG2* in 2010

	Model 1	**Model 2**	**Model 3**
	SC, SG	***SC, SG1***	***SC, SG2***
$\rho^{D} - \rho^{ND}$	0.4049***	0.4814***	0.9513***
	[4.6650]	[5.3121]	[5.2735]
$\rho^{D} - \rho^{C2}$	0.4831***	0.6339***	0.5416***
	[7.9626]	[10.3059]	[6.0383]
$\rho^{C2} - \rho^{ND}$	-0.0782	-0.1526	0.4096***
	[-1.1193]	[-1.9339]	[3.4480]

Table A1: Sample structure in 2018

	Whole sample	**Sub-sample with claim**	**DOA**	**non-DOA**
claim	0.0371			
SC	0.0090	0.2420	0.2695	0.2275
RG	0.3400	0.3614	0.4912	0.2931
AB	0.3500	0.3708	0.5039	0.3007
SG	0.2189	0.2697	0.3505	0.2272
D	0.2795	0.3452	1.0000	0.0000
female	0.5599	0.6515	0.6541	0.6502
age2025	0.0087	0.0057	0.0061	0.0055
age2530	0.0376	0.0401	0.0344	0.0430
age3060	0.7674	0.8107	0.7968	0.8180
carage0	0.0439	0.1970	0.2246	0.1825
carage1	0.0591	0.1702	0.1705	0.1701
carage2	0.0644	0.1477	0.1424	0.1504
carage3	0.0608	0.1132	0.1106	0.1146
carage4	0.0618	0.0993	0.1004	0.0987
veh_m	0.2800	0.3320	0.3372	0.3292
veh_l	0.1612	0.1881	0.1910	0.1866
sedan	0.9719	0.9951	0.9962	0.9945
lnprem	8.9993	9.0407	9.2992	8.9045
bonus	0.9370	0.7083	0.7316	0.6960
Observations	269475	10010	3455	6555

Table A2: Conditional dependence between SC and SG (year 2018)

	dealer		Non-dealer	
	SC	***SG***	***SC***	***SG***
RG	-0.1621*	3.9976***	-0.0552	4.1421***
	[0.0833]	[0.1734]	[0.0539]	[0.1275]
female	0.1163**	0.1735	0.1878***	0.1635*
	[0.0565]	[0.1131]	[0.0379]	[0.0839]
age2025	-0.4879	1.7509***	-0.3606	-0.2547
	[0.4042]	[0.6026]	[0.2516]	[0.7147]
age2530	-0.4798***	0.5993**	-0.6081***	0.1064
	[0.1719]	[0.3043]	[0.1090]	[0.2218]
age3060	-0.0701	0.2061	-0.0548	-0.0222
	[0.0716]	[0.1511]	[0.0509]	[0.1054]
tramak_n	-0.5044	1.2453**	0.4607**	-0.2498
	[0.5860]	[0.5731]	[0.2073]	[0.5385]
tramak_f	0.3979***	0.1127	0.5452***	0.7828***
	[0.1484]	[0.2855]	[0.0839]	[0.1915]
tramak_h	0.1188	0.2980	0.2119***	-0.3979**
	[0.1440]	[0.2660]	[0.0795]	[0.1850]
tramak_t	0.5698***	0.1843	0.3826***	0.3883***
	[0.0578]	[0.1162]	[0.0393]	[0.0891]
tramak_c	0.4150	-0.4958	-0.0831	0.3025
	[0.2818]	[0.6022]	[0.1984]	[0.4272]
carage0	0.2969***	-0.9596***	0.2489***	-0.6280***
	[0.0899]	[0.1991]	[0.0581]	[0.1419]
carage1	-0.1035	-0.9074***	-0.0309	-0.3448**
	[0.0878]	[0.1917]	[0.0576]	[0.1366]
carage2	-0.2374***	-0.1850	-0.0843	-0.3009**
	[0.0891]	[0.1909]	[0.0579]	[0.1320]
carage3	-0.2848***	-0.2969	-0.3812***	-0.5707***
	[0.0980]	[0.1936]	[0.0658]	[0.1408]
carage4	-0.2814***	-0.3862*	-0.2758***	-0.9843***
	[0.1003]	[0.2060]	[0.0684]	[0.1550]
veh_m	-0.0768	-0.0678	0.1282***	0.2059**
	[0.0593]	[0.1243]	[0.0407]	[0.0956]
veh_l	-0.3323***	-0.4531***	-0.1396***	0.3258***
	[0.0739]	[0.1601]	[0.0525]	[0.1124]

sedan	-0.5778	-1.3503*	-0.4488	-1.1315**
	[0.4758]	[0.6892]	[0.2802]	[0.5508]
lnprem	0.1095**	-0.0162	-0.0536*	-0.0823
	[0.0453]	[0.0916]	[0.0291]	[0.0592]
bonus	0.9318***	0.3404	0.7223***	-1.0135***
	[0.1128]	[0.2296]	[0.0790]	[0.1906]
ρ	0.5229**		0.9277***	
	[0.2575]		[0.2172]	

Standard errors in brackets; ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

The difference of conditional dependence between *SC* and *SG*:

$\rho^D - \rho^{ND}$ = -0.4048 (t = -1.7524; H_0: $\rho^D \leq \rho^{ND}$ cannot be rejected)

Table A3: Comparative manipulation ability of DOAs (years 2010 and 2018)

	First stage		Second stage	
	Est. coeff.	**P value**	**Est. coeff.**	**P value**
Intercept	-37.0773	<.0001	-0.7793	0.0027
SG			-0.4212	0.0899
$\widehat{SG}$			-0.1565	0.5312
dealer			0.00618	0.9478
y2018			-0.2399	0.0005
*SG*dealer*			1.8147	<.0001
*SG*y2018*			0.7457	0.0029
*dealer*y2018*			0.0443	0.6633
*SG*dealer*y2018*			-1.7265	<.0001
recoup	16.7495	0.8673	-0.1304	0.5139
female	0.3715	0.0198	0.0640	0.0322
age2025	-4.3641	0.9975	-0.4786	0.0182
age2530	0.0644	0.9055	-0.5400	<.0001
age3060	0.6748	0.0241	-0.0957	0.0182
tramak_n	-4.9743	0.9972	0.1025	0.575
tramak_f	0.6856	0.0085	0.0525	0.4566
tramak_h	-0.3143	0.3305	0.0618	0.3446
tramak_t	-0.3328	0.0447	0.2423	<.0001
tramak_c	0.3515	0.4887	-0.2892	0.0562
carage0	0.2408	0.2913	0.4741	<.0001
carage1	0.1337	0.5468	0.1807	<.0001
carage2	-0.2893	0.2825	0.1914	<.0001
carage3	-0.3436	0.2325	0.0823	0.1051
carage4	0.3167	0.1901	0.0314	0.5563
veh_m	-0.1722	0.3223	0.0300	0.3498
veh_l	-0.1807	0.3411	0.0333	0.3941
sedan	-1.4823	<.0001	-0.1291	0.3234
logprem	3.8113	<.0001	-0.0223	0.3687
bonus	-0.8332	0.0087	0.5239	<.0001
-2logL	12270.286		11655.898	

Table A4: Testing Hypothesis 2 for year 2018

	SG		**Type C**	
	Est. coeff.	**P value**	**Est. coeff.**	**P value**
Intercept	20.8341	0.0259	30.3130	0.0009
SC	-4.8308	0.0581	-9.0929	0.0006
first	-0.8095	0.5265	-3.7417	0.0036
*first*SC*	-0.7615	0.7789	1.8227	0.5221
female	-3.7137	<.0001	-0.7140	0.4280
age2025	-10.3178	0.0970	11.9403	0.0314
age2530	-3.5532	0.1216	3.4655	0.1534
age3060	-2.5520	0.0254	-0.9612	0.4385
tramak_n	-1.7228	0.7422	-7.9312	0.1331
tramak_f	-4.9029	0.0172	-9.3758	<.0001
tramak_h	-7.9695	<.0001	-11.4974	<.0001
tramak_t	-7.0401	<.0001	-11.1717	<.0001
tramak_c	-8.6608	0.0585	-10.9192	0.0065
carage0	2.7724	0.0447	3.4785	0.0142
carage1	5.1712	<.0001	4.1478	0.0029
carage2	3.7767	0.0039	5.8271	<.0001
carage3	2.5255	0.0692	5.5491	0.0003
carage4	2.3239	0.1084	4.1172	0.0100
veh_m	5.0363	<.0001	4.5882	<.0001
veh_l	9.7725	<.0001	15.1139	<.0001
sedan	6.2842	0.3710	7.3519	0.2293
logprem	-0.6929	0.2462	-1.5460	0.0395
bonus	2.1328	0.2304	-0.0887	0.9615
$Adj.R^2$	0.0773		0.0549	
Observations	3149		7530	