Knowledge about insurance products, adverse selection and moral hazard: Evidence From a randomised Experiment in Mongolia

Delger Enkhbayar*

February 17, 2015

[PRELIMINARY AND INCOMPLETE]

Abstract

Lack of knowledge about insurance products is common in developing countries. Individuals facing such demand frictions may choose products not well suited to their risk type. Particularly in settings where firms lack capacity to use risk adjusted pricing, firms can misinform consumers, such that the extent to which government should intervene and inform consumers is a relevant question. Given that, in this paper I study the extent to which lack of understanding in insurance products affects adverse selection and moral hazard in a car insurance market in Mongolia. Combining a randomised experiment with survey data on direct measures of knowledge about insurance coverage, I find that in our setting all adverse selection is driven by individuals who know their coverage at the point purchasing insurance. In other words, those who understand well their coverage at the choice stage choose insurance best suited to their risk type. These results are robust to looking at unclaimed accidents as well as history of accidents prior to the contract, suggesting learning through claiming does not fully explain these findings. Not knowing own risk type does not appear to be important in explaining variation in informational value. On the other hand, I find that informational gain is particularly high for buyers who have: 1) have higher income, more expensive cars and higher education, 2) are more risk-averse, and 3) bought insurance through sellers whom they know personally or through friends. There is no evidence of moral hazard in our setting.

*STICERD, Department of Economics, London School of Economics and Political Science (e-mail: d.enkhbayar@lse.ac.uk)

I Introduction

Asymmetric information offers one potential explanation for distortive insurance provision (Akerlof, 1970; Rothschild & Stiglitz 1976; Wilson 1977). An insurance contract specifies a premium and amount of coverage. Assuming individuals are well informed about their own risk type as well as alternative insurance products, they self-select into a contract which suits best to their risk type. Classic economic models of insurance predict a positive correlation between amount of coverage and exposure to risk through ex-ante adverse selection, with riskier individuals more willing to buy the high coverage. This correlation is further strengthened by presence of ex-post moral hazard through individuals with higher coverage taking less precautions to prevent accidents.

Despite this, individuals in developing economies often lack basic knowledge about insurance products or even what their purchased insurance contract covers against. Particularly in contexts where firms lack capacity in designing risk-adjusted pricing well, they may use lack of information or similar tools as a means to reduce adverse selection and moral hazard[1]. In such settings, it is important to understand the extent to which lack of information affects adverse selection and moral hazard. What are the characteristics of individuals who respond most to information, both at the insurance choice and at the claims stage? Finally, what are the welfare implications of policies aimed at providing information and advice to potential insurees?

We use a randomised experiment in vehicle insurance markets in Mongolia to separately identify adverse selection and moral hazard for different types of vehicle risks. In our setting a contract specifies a premium and coverage against different vehicle risks (collision, third party and theft). Coverage increases incrementally in the sense that the most basic insurance coverage is against collision, followed by one against collision and third party, and finally comprehensive insurance covering all three risks. To distinguish between adverse selection and moral hazard for different types of risk, some insurees are randomly offered a higher coverage at no additional cost. Adverse selection or selection on factors correlated with risk is identified by comparing risk measures between those who self-selected in different insurance coverages, but ended up with the same contract. On the other hand, ex-post moral hazard is identified by comparing risk measures between individuals who self-selected in the same insurance coverage but were assigned to different contracts by the randomisation.

[1]This applies to any strategy making it costly for individuals to acquire information about insurance, including complicated menu design, highly technical wording, information with multiple meanings, unclear instructions on the claims process et cetera.

Combining the above randomisation with extensive survey data collected one year after the experiment started, in which I ask individuals about details of their insurance contract and verify against their actual contract, I look at how adverse selection and moral hazard interact with how well an individual knows her own contract. If better informed individuals self-select in contracts more suitable to their risk level than uninformed individuals, policies informing consumers lead to a starker welfare trade-off between "better" choices and less distortive insurance provision due to asymmetric information. Finally, I examine potential sources of heterogeneity in gains from information. Certain features of our context further strengthen the relevance of the above argument. Firstly, the aforementioned trade-off is particularly relevant to insurance markets, where the cost of insurance provision depends on the risk pool of insurees. Secondly, insurance providers in developing countries often lack capacity in designing risk-adjusted pricing well. Finally, related to the second point, they often engage in strategies aimed at confusing/misinforming consumers, both at the choice and the claims stages.

We find evidence of adverse selection in both third party and theft risks, though theft risk is rare. For third party risk, adverse selection result only holds for individuals who know whether they are covered against third party risk. This is robust to changing the dependent variable to unclaimed risk measures as well as history of accidents prior to the contract, suggesting that our measure of information can be interpreted more broadly, including information at the point of purchasing insurance. In other words, rather than merely learning about insurance through claiming (after insurance purchase), individuals' choices also reflect ex-ante information. I also find that individuals gain more from information, that is make choices more coherent with risk type, if they have: 1) higher income, more expensive cars and higher education level, 2) higher risk aversion, and 3) bought insurance from a seller they know personally or through friends. Those with more driving experience/car usage do not profit more from information, suggesting information acquiral does not seem to depend on knowing own risk type. For theft risk, interaction with knowledge is harder to estimate due to a much smaller sample size as well as rarity of theft risk. Overall, I find no evidence of moral hazard for either risk.

Our research is related to a large empirical literature dedicated towards understanding the role of asymmetric information in explaining insurance market inefficiencies (for a comprehensive review, see Einav et al., 2010). This work owes much to the efforts of Chiappori & Salanié (1997), who describe a set of positive correlation tests for asymmetric information. In competitive markets, a significant positive correlation between coverage and ex post risk, conditional on all information used by the firm in pricing, would indicate presence of asymmetric information: either consumers had prior information about their exposure risk (adverse selection), or insurees with higher coverage took less care (moral hazard), or both. In vehicle insurance markets existing empirical results using this conditional correlation approach are mixed. Some find evidence of asymmetric information (Dahlby, 1983, 1992; Puelz & Snow, 1994; Richaudeau, 1999; Cohen, 2005; Kim et al, 2009), while others do not (Chiappori & Salanié, 2000; Dionne et al, 2001, 2006).

While positive correlation tests provide many valuable insights into the presence of asymmetric information in many markets, they face the difficulty in differentiating between adverse selection and moral hazard. Clearly, for a researcher evaluating effectiveness of policies, such as making comprehensive insurance more accessible, this distinction is important. On this Abbring et al. (2003) exploit panel data on insurance choices and claims to distinguish moral hazard from dynamic selection on unobservables. Under moral hazard, experience rating implies negative occurence dependence, since past claims increase the cost of filing for an additional claim. With this methodology they find no evidence of moral hazard. A second set of papers looks at experimental variation in "actual" versus "selected" contracts, for instance Karlan & Zinman (2009) in consumer credit, Finkelstein et al. (2012) in health insurance, Gunnsteinsson (2013) in crop insurance. I use a randomised experiment for identification, so in this sense, our methodology is similar to this literature, however, our focus is on interaction of adverse selection and moral hazard with the degree of knowledge about insurance.

Recent empirical work is also motivated to move towards more structured approach, by building empirical models that incorporate rich heterogeneity in consumer preferences as well as the heterogeneity in risk emphasized in classic theoretical contributions. This is well illustrated by Finkelstein & McGarry (2006), who find evidence of adverse selection in long term care insurance once they control for risk aversion, proxied by the extent to which individuals take preventative health care measures. While highlighting the potential sensitivity of reduced-form estimates to model specification, this literature tends to abstract away from moral hazard and focuses on selection on preferences or other characteristics, such as risk aversion (de Meza & Webb 2001, Finkelstein and McGarry 2006, Cohen and Einav 2007), cognitive ability (Fang et al. 2008), desire for wealth after death (Einav, Finkelstein, and Schrimpf 2010) or behavioural biases (Spinnewijn 2012, Barseghyan et al. 2011, 2012).

Empirical literature on interaction of adverse selection and/or moral hazard with information about insurance coverage is limited, presumably due difficulty in directly measuring information and linking this to administrative data. Within the above literature on multidimensional asymmetric information, our work is most related to Fang et al. (2008), who show that cognitive ability is a prominent source of advantageous selection in Medigap, a health insurance product offered to the elderly by private companies to fill in the "gaps" present in Medicare. Initially they find a negative correlation between Medigap coverage and total medical expenditure (controlling for the determinants of price), suggesting presence of multidimensional asymmetric information. After controlling for direct health measures the correlation turns positive, with the implication that healthier individuals have both lower medical expenditure and are more likely to enrol in Medigap (advantageous selection). They gradually add controls of potential sources of advantageous selection to the regression of Medigap coverage on predicted medical expenditure (controlling for risk aversion and the determinants of price) and find that estimated coefficient changes substantially as cognitive ability is added. With this, on average the elderly citizens who have higher cognitive ability are both more likely to buy Medigap and are healthier.

Our paper is also related to recent literature on welfare analysis of counterfactual menu designs in presence of informational frictions. While insightful, this literature focuses on structural estimation and abstracts away from private information and moral hazard. For instance, Handel & Kolstad (2013) uses a direct measure of knowledge of alternative insurance products to estimate a structural model of insurance choice with informational frictions and hassle costs. They find that these additional frictions affect individual choices, and estimated risk aversion, which has important welfare implications, is much higher, if frictions are omitted from the analysis. Welfare loss from risk exposure, when frictions are not taken into account, is more than double when they are. Their paper does not identify private information and moral hazard, and focuses on the impact of informational frictions and hassle costs on choice biases.

Handel (2014) identifies the extent to which inertia worsens individual choices and evaluates the extent to which nudging affects adverse selection and welfare. They find that nudging substantially exacerbates adverse selection, and reduction in welfare is double than the existing welfare loss due to adverse selection. They examine how specifically nudging in health insurance affects welfare within a framework of modelling choices with inertia, while our paper on the other hand focuses on identification of adverse selection and moral hazard using a randomised experiment and how these interact with direct measures of information. Furthermore, I discuss potential sources of information acquiral using additional data on buyer-seller relationships and other individual-specific characteristics.

The rest of the paper proceeds as follows. Section II describes the institutional background with an emphasis on informational environment both at the firm and at the individual level and how these features affect methodology choice as well as interpretations of our findings. Section III presents experimental design/implementation and a detailed description of administrative and survey data, in particular direct measures of information about insurance. Section IV presents the identification strategy and preliminary results, and I show both evidence of adverse selection and how it is exacerbated by availability of information. Section V presents preliminary discussion on potential sources of variation in information, focusing on characteristics of individuals who profit most from information, and how both access to and acquiral of information are relevant in explaining our results. Section VI concludes.

II Context and menu design

Here I describe the features of the informational environment in which firms and consumers interact most relevant to our research design and interpretations of the results in section IV. I start with market overview and insurance sales channels. Next I discuss how insurance firms (and regulators) in Mongolia often lack capacity and know-how in monitoring, collecting and analysing data, and finally designing risk-adjusted pricing for almost all insurance products. I also present anecdotal evidence for insurance providers' often engaging in strategies to confuse consumers both at the choice stage and the claims procedure. Finally, I present the experimental menu design.

A The market overview and sales channels

The market for private insurance is relatively small in Mongolia. According to the Financial Regulatory Commission's 2011 report, less than 10% of cars had some type of insurance before Third Party Liability insurance (TPL) became compulsory at the end of 2011. Top five insurance firms, out of 18 in total, earn on average 80% of the car insurance market revenue[2]. Mongol Daatgal LLC. (MD from here on), where I implement the experiment, is the largest accounting for 30% of the car insurance market revenue. Firms offer somewhat different car insurance products and often information about pricing is not publicly available. Only 30% of our survey respondents compare MD car insurance with competitors' products. These suggest some type of imperfect competition.

By law insurance coverage only starts after the buyer physically signs the contract, thus forbidding sales through internet. Hence, all contracts are

[2] 17 out 18 insurance firms provide non-life insurance, with one firm specialising in life insurance. Yearly data on firm-level revenue are obtained from www.frc.mn, here I report the average over 2009-2012.

paper-based. There are three types of insurance suppliers: 1) branch managers and agents, 2) brokers and 3) banks. Branch managers sell at the branch and receive a fixed salary as well as a bonus of 1.8% of all generated income. Agents often operate outside the branch without access to internet/insurance program, for instance at gas stations, markets, vehicle registration offices et cetera. They receive a bonus of 15% of generated income and no salary (see Appendix A for more details on their incentive mechanism).

In 2011 banks were permitted to sell car insurance, creating a "parallel" market with a difference that bank insurees are primarily borrowers insuring their collateral. Both banks and brokers contract non-exclusively with the insurance firm, so bank borrowers typically face a larger choice set than branch insurees. Although most contracts are sold through banks and brokers, at least for first-party car insurance, I decided against implementing the experiment with them for two reasons. Firstly, banks recommend insurees towards a particular coverage in a non-systematic way, so buyer "choices" do not necessarily reflect their risk type. Secondly, MD was not willing to design an identical insurance product for banks to remain more competitive, and finally, there were concerns on how non-branch sellers would handle the randomisation process.

B Informational environment

MD has been using the same insurance program for at least a decade, failing to record data on most individual- and vehicle-specific characteristics known to be important in insurance pricing, for instance age, gender, vehicle type, engine size et cetera.[3] The only variables the program records are contract identifier, branch, insurance start and end date, car valuation and total premium paid. For first-party car insurance, before the experiment the firm offered a complicated menu design, while also allowing sellers to flexibly set prices within a range. Given that the program only recorded contract-specific premium paid, it was not feasible to identify contract-specific amount of coverage or the type of insurance sold from the firm's program data. Also, MD did not record systematically additional data from paper-based contracts, so it is rather puzzling how prices are set within this framework. Moreover, data on the history of claims are only centralised for TPL insurance since it became compulsory in 2011. For other insurance products not only the claims data are not available centrally, but MD lacks capacity in recording history of claims for its own insurees. As of 2013 no insurance firm has yet introduced a bonus malus system

[3]At least until the end of our experiment, the firm continued on this insurance program. In 2013 it started developing a new insurance program.

for first-party car insurance in Mongolia[4]. There is also little regulation in this market.

Given the above limitations, in our context firms lack capacity in designing risk-adjusted pricing well. Limited data collection and analysis as well as difficulties in implementing improvements are further exacerbated by lack of technology and know-how. Lack of competition, as mentioned before, may reduce incentives to innovate.

In this environment the firm might want to use other tools to reduce adverse selection and moral hazard. One mechanism is to increase uncertainty on details of the contract, both when insurees are choosing coverage and at the claims stage. There is anecdotal evidence on the degree of such misinformation. From discussions with sellers and directors, it was clear that many were in favour of a complicated menu design with more flexibility in bundling and pricing risks. While some claimed that this helped sales, others suggested that its popularity is due to more margin for confusing consumers. The experimental product design is simpler, despite this only 20% of survey respondents understood fully all the features of their insurance coverage, with the rest either not knowing or misinterpreting their actual coverage[5]. Survey interviews also suggest general dissatisfaction with the claims procedure, with insurees often misinformed about the claims application process as well as how payouts are determined (see Appendix C for brief description of the claims procedure).

C Menu design

First-party vehicle insurance implies damages only to the insured vehicle are covered by the firm. As mentioned in part B, before the experiment, the firm offered a complicated product design with some flexibility for sellers to adjust prices. MD's business development team with the actuary had been previously working on a much simpler product design and I decided to keep that design for the experiment in Figure II.1.

Now I provide the firm's definition of each risk included in the experimental product design. Vehicle accident is defined by the insurance company as any solid material changing the original state of the car. This includes collision, hitting a post or garage door, being hit by a stone (which is a common risk on dirtroads) et cetera. Natural disaster includes hail, thunder and any other acts of God. Third party (deliberate) damage is any intentional damage to the vehicle by a third party, for instance vehicle scratched or drawn on while

[4]Some type of bonus malus scheme, not in its usual sense, exists at MD. See Appendix B for more details.

[5]Despite this, around 70-80% of survey respondents knew some features of their contract.

being parked, hit-and-run et cetera. Typically, the firm categorises a risk to be a third party risk, if the vehicle is parked and the faulty party cannot be traced[6]. Drivers' personal accident pays out 5 million MNT for 70% or higher loss of employability or death as a result of a car accident. There is a separate personal accident cover offered by the firm, with a higher protection level. Water leakage in the garage, fire/explosion, theft/robbery are obvious.

Figure II.1 describes the different plans available in the experimental product design. Amount of coverage is determined by a bundle of risks and a co-insurance rate, the proportion of losses covered by the insuree herself. Total premium paid simply depends on amount of coverage and car valuation. For example, product 1 covers 90% of losses due to "Vehicle accident" and "Natural disaster" at a premium of 0.8% of car valuation. I do not allow sellers to flexibly set prices to keep the same menu design for all consumers. Due to the existing bonus structure, lack of data and lack of underwriting skills, sellers whether they are allowed to flexibly set prices, do not price contingent on risk, with a further implication that buyers are never refused insurance. Furthermore, the firm does not customise the premium rates according to individual-specific characteristics, other than car valuation.

[6] These definitions imply that for third party risk the firm cannot retrieve claims from the party at fault (no subrogation), while for collision it can (subrogation). Some of the price differential may be explained by this.

	RISKS	PRODUCT 1	PRODUCT 2		PRODUCT 3	
RISK 1	Vehicle accident (in motion)	√	√		√	
	Natural disaster	√	√		√	
RISK 2	Water leakage in the garage		√		√	
	Third-party (while parked)		√		√	
RISK 3	Fire/explosion				√	
	Theft/robbery				√	
	Drivers' personal accident				√	
	PRICING					
	Premium (% of vehicle valuation)	0.8%	1.2%	2.0%	3.0%	3.8%
	Co-insurance rate (%)	10%	10%	0%	10%	0%

Figure II.1: Product design

III Methodology

A Experimental design and implementation

In order to identify adverse selection and moral hazard separately I insure randomly selected individuals against an additional source of risk for a first-party vehicle insurance product. Some individuals who self-selected in "Vehicle accident" cover (product 1) are randomly insured against "Vehicle accident" and "Third party" (product 2 with 10% co-insurance rate) for free. Similarly, a random set of individuals, who self-selected into "Vehicle accident" and "Third party" cover (product 2 with 10% co-insurance rate), are insured additionally against "Theft/robbery" (product 3 with 10% co-insurance rate) at no additional cost[7][8]. I also randomly assign a subset of insurees who self-selected into a coverage with a high co-insurance rate (product 2 with 10% co-insurance rate) to one with a lower co-insurance rate (product 2 with no co-insurance rate) at no additional cost. The arrows in II.1 indicate the three types of randomisations I implemented.

RCT was implemented by Mongol Daatgal LLC., the largest insurance provider in Mongolia, for exactly one year from July 2013. As mentioned in the context section, sales through branches are the sole channels through which insurance is fully voluntary, hence the randomisation was implemented at all the seven branches of the firm in Ulaanbaatar, with 51 insurance managers and 12 active agents. While insurance sales lasted until July 2014, claims data will continue to be collected until July 2015. So, all our results are based on six months' worth of claims data for each one-year contract on average. Only at the end of July 2015 results will reflect full data on claims.

Due to all contracts being paper-based and significant off-site selling, there were concerns on the possibility that sellers might try to find out the randomisation outcome before finalising the contract, biasing the randomisation. To stop this from happening, the randomisation was implemented by a mobile messaging system with sellers asking buyers to send SMS from their mobile phone right after the contract is signed. The server would receive the message, containing some contract information, and respond to both the buyer and the seller with the randomisation outcome. If the server sent the randomisation outcome and the contract was not finalised, there was a punishment to the seller of one month's minimum wage. This combined with training throughout the project on interacting with the mobile program and daily monitoring through the messaging system ensured sellers had the incentive to guaran-

[7]There are additional risks included in the covers but they are rare. In our experiment there are no claims for risks other than "Vehicle accident", "Third party" and "Theft".

[8]These are similar to offering free additional insurance to a random set of uninsured individuals.

tee the contract was sealed before the randomisation outcome was found out (see Appendix D for brief description of the instructions/manual sent out to sellers).

B Data description

In July 2014 when the project ended, I audited all bills that were distributed to branches. Out of 2565 bills, 1434 were valid contracts, the rest were either invalid or unused. Only 39 messages were sent for the invalid/unused contracts, which suggests the monitoring worked. Six contracts are dropped from the analysis as they have not yet started by July 2014. The message scheme allowed me to detect 36 contracts suspected of biasing the randomisation[9]. I will exclude these contracts from the analysis, though for all regressions coefficient estimates do not vary significantly, if they are included.

Balance tables to check the validity of the randomisation are presented in Appendix II. Variables used in these tables are individual- and vehicle-specific variables handwritten on the contracts. Individual-specific variables are insuree gender, age (in years) and dummy variables for the district of residence (I aggregate residence outside of Ulaanbaatar as "Other"), and vehicle-specific variables are vehicle age (in years), valuation (in Mongolian Tugriks MNT)[10], dummy variables for the vehicle brand, a dummy for whether vehicle has a lighter or a darker colour. Additional variables, such as seller-specific characteristics (gender, age, length of employment at MD, a dummy for whether the seller is an agent or a manager) as well as the number of days for which the contract was valid until 15/07/2014 when the data collection started. Table II.1 shows the differences between those who chose product 1 and were assigned to product 2 with 10% coinsurance rate *versus* those who remained in product. Similarly, Table II.2 presents the relevant results for those who chose product 2 with 10% coinsurance rate and were assigned to product 3 with 10% coinsurance rate *versus* those who remained in product 2. Finally, Table II.3 shows the results for those who chose product 2 with 10% coinsurance rate and were assigned to product 2 with no coinsurance rate *versus* those who remained in product 2 with 10% coinsurance rate. Overall, there is a decent balance on 18 out of 21 pre-randomisation outcomes. Note that Days insured and whether the seller is an agent are significantly different from 0 on two of the three randomisations. This is due to seasonality of sales and

[9]For 22 contracts, messages were sent at least one day before the contract has started, for 3 the randomisation was incorrectly done, and the rest were related to the above 39 messages for invalid/unused contracts.

[10]1 GBP $\approx$ 3000 MNT according to www.mongolbank.mn as of February 2015.

branches with high volume sales (and managers rather than agents) randomly receiving more contracts with upgrades[11].

1 Administrative data

I collected, documented and entered data for 1434 contracts and 411 claims materials. Contracts had a standard format, while claims materials did not, ranging from five to forty pages, hence common overlapping documents, including insurance hotline reports, claims application forms and claims department reports, were identified and the data were entered manually. For some claims materials even these were missing, so I filled in the gaps through enquiring with claims managers, hotline employees or police reports.

Both contracts and claims materials were handwritten, and since the company did not collect any data from the contracts themselves, I took photos and did manual double entry of data[12]. Variables collected in the contracts data are individual-specific (insuree gender, age and residence), vehicle-specific (manufacture year, valuation, brand, colour) and contract-specific (insurance coverage, start and end date, outcome of the randomisation stated on the contract, a dummy for whether a signature of the buyer is present and dummy variables for the seller) variables. I collected a vast set of variables from the claims materials, including risk type/category, claim size, payout, accident date, claim decision date, damage valuation, car valuation at the time of accident, dummy variables for the damage evaluator et cetera.

Seller-specific characteristics, including gender, age, length of employment at MD, insurance sales experience as well as whether the seller is an agent or a manager, were obtained from human resources records.

2 Survey data

I interviewed 552 insurees (the respondents, group A), after having tried to contact all insurees[13]. Some insurees had more than one vehicle insured, so the total number of contracts for respondents reached 570, 42% response rate. Insurees from 339 contracts refused to participate in the survey (group B),

[11] Distribution of contracts is random and proceeds as follows. Branches request a specific number of contracts throughout the year and the central office sends contracts ordered by a numeric contract identifier. Within branches distribution to managers and agents is also random.

[12] I checked data for some variables in the data entered manually against those that were recorded in the firm's program data. Also, a random sample of the entered data were checked against the original contracts. During these checks I found no errors.

[13] 43 and 25 out of 1434 contracts stated residences outside of Ulaanbaatar or were of foreigners, respectively, so they were not contacted.

either because they were busy or for unrecorded reasons. For the remaining 519 contracts, I was either unable to reach the insuree or they could not participate within the survey time frame (group C). Tables III.1, III.2 and III.3 are balance tables for group A *versus* group C , group A *versus* group B and group A *versus* groups B and C, respectively. Overall, there is a balance on 1) the treatment variable "Upgrade" taking value 1 if an individual is upgraded to a higher coverage and 0 otherwise, 2) risk outcomes (overall and by risk category), and 3) a range of other individual- and vehicle-specific controls from independently collected data. Age and car valuation enter significantly different from 0 on all balance tables. Differences for car valuation is driven by extreme values, so if I eliminate top 1% of car valuations, the difference becomes insignificant (for example, see III.4 for Group A vs Group B, C comparison). Age on the other hand is not driven by extreme values and median survey respondent is older than median non-respondent. However, I include age as a control variable in all my regressions.

Our interviewers are not given in advance any information about the insuree, except the full name, telephone number and home address recorded on the contracts. We told the interviewees in our initial calls that their information was shared by MD as part of a collaborative work to understand driving and insurance experience in Ulaanbaatar. I also told them that surveys will be undertaken independently by a group of researchers with external funding and no information provided during the surveys will be shared with a third party, including MD. Interviewers were instructed to inform the respondents of this again just before the survey began. This was in attempt to obtain a high response rate (and explain how I obtained their personal information) as well as tackle problems of underestimating riskiness and risky driving behaviour. Either way, around 1500 risks were recorded as part of three year accident history for all respondents, so at least quantitatively revealing this information did not seem to be a major issue. The interviewers each ran about four interviews a day, lasting 45 minutes on average, with the time in between spent repeatedly contacting insurees to schedule or re-schedule interviews. I collected information on the interview process, such as the length of the audio proxying for interview duration (sometimes audio recording stopped short before the interview and interviews had to be rescheduled), the time of the day and the day of the week, as well as the identity of the interviewer, in an attempt to use these as controls to improve precision of our estimates.

When the insuree was unavailable for an interviewee, he/she suggested we interview a household member, there were 21 such cases. In 20 of 21 of these cases, the respondents drove the insured vehicle and 16 bought the car insurance together with the insuree. On the other hand, there were rare cases of interviewing the insurees and realising the information they provided did not fully fit our purpose. For example, there was a case of a respondent who bought insurance for her relative, while herself not having a driving licence, while some others were not main drivers of the vehicle, with the implication that some variables, such as vehicle usage, were subject to measurement error. Additionally, since interviewers were not given any information on the car insured by MD, in 65 cases the vehicle information provided by insurees did not match that of the insured vehicle, mainly due to respondents having sold or lent the car and hence not reporting it.

The survey data can be broadly categorised into six broad areas:

- Individual-specific characteristics, for instance education, employment and direct and indirect measures of income (all category variables).
- Vehicle-specific characteristics not recorded by the firm, for instance vehicle usage (average distance travelled per day, a dummy variable for work driving, a dummy for whether the vehicle is used by drivers other than the insuree), engine size, vehicle ownership, investment into the vehicle.
- Different measures of individual-specific riskiness and/or effort level, including driving experience, where they park their vehicles at home and at work, a dummy for theft signal, three year accident history with detailed information about each accident[14] and evaluation of risky activities while driving.
- Risk aversion proxy, asking insurees to consider a binary lottery choice between a safe option yielding a certain amount and a risky option, in a sequence each time increasing the safe option by a certain amount. I also ask insurees whether they have bought in the past other types of insurance, which could serve as a risk aversion proxy.
- Different measures of self confidence in driving, including evaluation of own driving skills and subjective probability of accident in Ulaanbaatar by risk category.
- Measures of knowledge of own insurance coverage. Questions to capture knowledge included: 1) How well did you examine the alternative coverages?, 2) In your current contract, are you covered against "Collision"/"Third party"/"Theft" risk?, 3) What is the co-insurance rate if any? "Don't know" is optional.

[14]In order to classify accidents into risks in the experimental product design, during the interview, we explained the survey respondents each of these risks with both a formal definition and a list of common examples. Throughout the survey we did not mention the actual product design or the randomisation, but sometimes the insuree would ask during questions on information about coverage.

- Trust in insurance: whether they trust MD, whether they would recommend MD vehicle insurance to friends/family, whether they had any conflict with insurance companies and how it was resolved, three year history of claims applications and their outcomes.

I also surveyed sellers on whether they knew personally or through friends the buyer for each contract. For around 27% of contracts, sellers knew personally or through friends the buyer.

IV Empirical results

A Identification strategy

Underlying theoretical framework is one with individuals differing in riskiness across multiple risks $R = \{$"Collision", "Third party", "Theft"$\}$ as well as preferences. The methodology is similar to that in Karlan & Zinman (2009).

The presence of adverse selection predicts that, given the final amount of coverage insurees have, the group of insurees who self-selected into a lower coverage have lower expected risk than those who self-selected into a higher coverage. To capture this I run the following regression:

$$Risk_{i,R} = \alpha_{i,R} + \beta_{0R} Choosing\ lower\ ins_i + \beta_{1R} X_i + \epsilon_{i,R} \tag{1}$$

on the sample of individuals who chose insurance only differing in coverage of R but ended up in the same contract covering R. Both end up in the same coverage, so holding moral hazard constant, I identify selection. For instance, to identify adverse selection in "Third party" I compare risk measures between those who chose product 1 and were insured additionally by the randomisation against "Third party" versus those who chose product 2 (with 10% co-insurance rate). Both after the randomisation have exactly same amount of coverage, but one group initially chose to not get insured against "Third party". The dependent variable *Risk* are different measures of riskiness in R and the key explanatory variable *Choosing lower ins* is a dummy taking value 1 if the individual self-selected into insurance excluding cover against risk R or 0 if the individual self-selected into insurance including cover against R. X are individual-, vehicle- or contract-specific control variables. A significant negative coefficient $\beta_0 R$ would suggest presence of adverse selection, while positive significant coefficient $\beta_0 R$ would suggest multidimensional selection.

I run the following regression to estimate moral hazard for risk R:

$$Risk_{i,R} = \tilde{\alpha}_{i,R} + \tilde{\beta}_{0R} Upgrade_i + \tilde{\beta}_{1R} X_i + \tilde{\epsilon}_{i,R} \tag{2}$$

on the sample of individuals who self-selected in the same coverage excluding R but some were assigned to a higher coverage including R. Holding selection constant, this identifies the effect of a higher coverage on riskiness. For example, to estimate the extent of moral hazard for "Third party", I compare those who self-selected in product 1 and remained in product 1 versus those who self-selected in product 1 and were randomly upgraded to product 2 (with 10% co-insurance rate). Here it is important to note that we might not capture moral hazard simply by looking at claims data, as those who remained in lower coverage may still incur accidents while not being able to claim losses through insurance. Accident data is only available in the survey, so we run this regression only on survey respondents. The dependent variable *Risk* are different measures of riskiness in R and our key explanatory variable *Upgrade* is a dummy taking value 1 if the individual was upgraded to insurance against risk R or 0 if the individual stayed in the same coverage. Again X are individual-, vehicle- or contract-specific control variables. A significant positive coefficient $\tilde{\beta}_0 R$ would suggest presence of moral hazard.

In order to identify heterogeneity of adverse selection (if it exists) across individuals differing in knowledge about their coverage, I run the following regression:

$$\begin{aligned} Risk_{i,R} = \hat{\alpha}_{i,R} + \hat{\beta}_{0R} \textit{Choosing lower ins}_i + \hat{\beta}_{1R} Knowledge_{i,R} + \\ + \hat{\beta}_{2R} \textit{Choosing lower ins}_i * Knowledge_{i,R} + \hat{\gamma}_R X_i + \hat{\epsilon}_{i,R} \end{aligned} \tag{3}$$

where the sample, the dependent variable *Risk* and covariates X are the same as in regression (1) and *Knowledge* is a dummy taking value 1 if the individual knows she is covered against R or 0 if she is misinformed about it. All insurees in this sample are insured against R but only 70% know that they are. We expect $\hat{\beta}_{2R}$ to be negative if informed individuals choose coverage more coherent with their risk than uninformed individuals.

Similarly, if we expect moral hazard only for individuals who know or remember their contracts after purchasing insurance, $\bar{\beta}_{2R}$ should be positive and significant :

$$\begin{aligned} Risk_{i,R} = \bar{\alpha}_{i,R} + \bar{\beta}_{0R} Upgrade_i + \bar{\beta}_{1R} Knowledge_{i,R} + \\ + \bar{\beta}_{2R} Upgrade_i * Knowledge_{i,R} + \bar{\gamma}_R X_i + \bar{\epsilon}_{i,R} \end{aligned} \tag{4}$$

where the sample, the dependent variable *Risk* and covariates X are the same as in regression (2) and *Upgrade* is a dummy taking value 1 if the individual knows she is covered against R or 0 if she is misinformed about it.

All regression results are presented in Appendices IV-VII. Depending on parameter of interest, the dependent variable will be from either administrative claims data or accidents data from the survey. If only administrative data is used, the dependent variables are claim frequency, probability and size. If survey data is used, the dependent variable can be history of accidents before the current contract (both unclaimed and claimed) and unclaimed accidents during the contract. I can use other measures of risk, for instance whether individuals park their vehicles in secure areas, tolerance of risky driving behaviour etc. Explanatory variables include log of days insured[15], gender, age, log of vehicle valuation, vehicle age, a dummy for vehicles of lighter colour as well as residence and brand dummy variables. I also include sellers grouped according to their experience (less than or equal to 3 years at MD) and whether or not they are agents. Additional explanatory variables could include insuree income, education and employment type, dummies for interviewers, the day of the week and the time of day the interview was conducted and the length of the interview.

B Adverse selection and moral hazard

Regressions for identification of adverse selection and moral hazard for third party and theft are in Appendix IV. For all regressions if the dependent variable is accident frequency or accident probability, then I run poisson or logit regressions, respectively, and present average partial effects. For loss size, I take $log(loss\ size + 1)$ and use least squares regression with robust standard errors. Additionally, given that on average, I have six months of claims data for each one-year contract, I expect that the regressions with administrative claims data sometimes do not have statistical power, particularly for a rare risk such as theft. Therefore, for theft I use as the dependent variable (i) claimed and unclaimed accidents during the contract period, (ii) history of accidents before the contract starts, and iii) three year history of accidents, merging i) and ii). Given that history of accidents prior to the contract has a high correlation with accidents during the contract period I expect these to be appropriate measures of riskiness.

Table IV.1 presents adverse selection results for third party risk using claims during the experiment as the dependent variable. It shows that given the same ex-post moral hazard, individuals who self-selected into coverage against "Vehicle accident" only are less risky in third party risk than individuals who self-selected into coverage against both "Vehicle accident" and "Third party".

[15] As data collection on claims is still in process, I need to control for length of time for which the contract is valid.

For the latter, average third party risk frequency and probability are 0.10 and 0.09. The point estimates in the first row are all negative and statistically significant. When all controls potentially correlated with riskiness are accounted for, the point estimates are slightly higher, while precision remains about the same. Columns 1-4 show that those who did not choose "Third party" insurance have at least 40% lower third party risk frequency and probability than those who chose "Third party" cover. Columns 5-6 show that those who chose lower coverage claimed around 50% less than those who chose higher coverage. These estimates are in magnitude very large, with the implication that adverse selection is both quantitatively and statistically significant in third party risk. Table IV.2 shows that adverse selection is stronger if 23 contracts suspected of tampering with the randomisation are eliminated (described in more detail in *Data description*).

Tables IV.3, IV.4 present adverse selection results for theft risk using claims during the experiment as the dependent variable. The coefficients on *Choosing lower ins* in the first row are positive and statistically significant in 4 out of 6 regressions, suggesting presence of advantageous selection. Those who self-selected in coverage against "Vehicle accident", "Third party" and "Theft" did not report theft, while those who self-selected in lower coverage against "Vehicle accident" and "Third party" reported two thefts. As claims for theft are rare, I run the same regressions but on the survey respondents, changing the dependent variable to: 1) accidents during the contract, 2) history of accidents prior to the contract and 3) three-year history of accidents. These are presented in Tables IV.5, IV.6, IV.7, respectively. We expect three-year history of theft to be a more accurate measure of riskiness in theft, and for this the coefficient on *Choosing lower ins* is significantly lower than 0 in terms of both theft frequency and probability. Average frequency and probability for those who chose higher coverage are 0.6 and 0.45, respectively. So, the point estimates in Table IV.7 suggest that those who self-selected in lower coverage are at least 50% less risky in theft than those who self-selected in higher coverage[16]. Columns 2, 4 and 6 show that controlling for individual- and vehicle-specific characteristics reduces precision for theft frequency, but not as much for probability and loss size. These results suggest that for a rare event such as theft we can only capture adverse selection, if we focus our attention to a long-run history of theft. The results are as with third party risk, both quantitatively and statistically significant.

Tables IV.9 and IV.10 show there is no evidence of adverse selection for

[16]Regressions should be re-run taking into account of the small sample size. For instance, instead of running a regular logistic/poisson regression, we should run an exact logistic/poisson regression as then estimates do not depend on asymptotic results.

co-insurance rate, using claims for both "Vehicle accident" and "Third party" as the dependent variable. Those who chose a lower co-insurance rate are not riskier than those who self-selected in higher co-insurance rate. Both are covered against "Vehicle accident" and "Third party", but those who chose higher co-insurance rate have to pay out of their pocket 10% of the loss. None of the point estimates in the first row are significant and the results are robust to including controls, as well as excluding contracts potentially biasing the randomisation.

Tables IV.11- IV.14 present moral hazard results for "Third party" and "Theft" using accidents during the experiment as the dependent variable. The coefficient on *Upgrade* is not significantly different from 0 for all regressions, except for loss size in "Third party" with the implication that those who were upgraded from insurance against "Vehicle accident" to insurance against "Vehicle accident" and "Third party" have at least 70% higher losses in third party risk. This is driven by (non-zero) claims and reported losses of zero for unclaimed accidents. This could be explained by: 1) presence of fraud with insurees covered against "Third party" claiming for more than actual loss size, 2) insurees with "Third party" coverage taking less precaution against more serious third party risks, for instance by parking in less secure spots, and 3) survey respondents with lower coverage under-reporting loss sizes. All are reasonable explanations for now, but later I will try to disentangle between them. For theft, the point estimates are positive but not statistically significant.

Tables IV.15 and IV.16 present moral hazard results co-insurance rate using administrative claims data as the dependent variable, while I use accidents during the experiment instead in Table IV.17. Since all insurees in the sample are allowed to claim for both "Vehicle accident" and "Third party", both claims and accidents can be used to capture moral hazard. The point estimates in the first row on *Upgrade* are not significantly different from zero, implying that there is no evidence of moral hazard.

C Interactions with knowledge

Adverse selection and moral hazard results are summarised in Table IV.18. Given that there is evidence of adverse selection in both third party and theft risks, I will run regressions (3) to understand how these results interact with information about own insurance coverage. Since there is no robust evidence of moral hazard in insurance against "Third party", "Theft" or lower co-insurance rate, regressions (4) will not be presented here. Regressions for interaction of adverse selection for third party and theft with knowledge are in Appendix V.

1 Adverse selection

First, I run regression (1) on the sample of survey respondents, using administrative claims data as the dependent variable. The results are presented in Table V.1, and we see that adverse selection is no longer significant, for instance in Column 3 point estimate on *Choosing lower ins* is -0.025 instead of -0.050 in the same location in Table IV.2. Later we will simulate 1000 random samples and look at the proportion of samples yielding point estimates higher or equal to those in Table V.1. If this probability is higher than 10%, then we can conclude that our survey is as representative as a survey of randomly selected individuals, at least in terms of adverse selection results. On the other hand, if we look at three year history of accidents in Table V.2, then there is evidence of adverse selection in third party in terms of frequency and probability.

Tables V.3, V.5 and V.7 present the results using as the dependent variable 1) claims during the contract, 2) history of accident prior to the contract, and 3) unclaimed accidents **during** the contract, respectively. The coefficient on *Choosing lower ins*Knowledge* are negative and statistically significant for frequency and probability of accidents. The coefficient captures correlation between knowledge about own coverage and “better” choices, so results imply that adverse selection is stronger for those who are better informed about their insurance coverage.

We can look at adverse selection for informed group and uninformed group separately in Tables V.4, V.6 and V.8, using the same dependent variables as above. In Table V.4, adverse selection is not robust among informed individuals while there is no evidence of adverse selection for uninformed insurees[17]. From Tables V.6 and V.8 we see that adverse selection holds only for informed individuals and there is no evidence of selection based on riskiness for uninformed group.

For theft, regressions need to be re-run accounting for small sample size. I ran exact logistic and poisson regressions for probability and frequency of theft accidents over three years and the coefficient on interaction of adverse selection and information is positive but statistically insignificant (not presented here). The results presented in Tables V.10 do not take into account small sample size, and the interaction term is positive and significant, suggesting that adverse selection is higher for uninformed individuals. In other words, uninformed individuals make choices more coherent with their riskiness than informed individuals. The standard errors are biased downwards if we run regular poisson and logit regressions, especially given that there are only 18

[17] For the informed individuals, the p-values are close to 0.10 for all point estimates. Precision may improve with more complete claims data.

uninformed individuals[18].

V Discussion

Using a randomised experiment, I identify adverse selection and moral hazard separately for "Third party" and "Theft" and coverage with a lower co-insurance rate. Overall, I find evidence of adverse selection for "Third party" and "Theft", and no evidence of adverse selection into coverage with a lower co-insurance rate. I find evidence of moral hazard only in loss size in "Third party", and no evidence of moral hazard for "Theft" and coverage with lower co-insurance rate. I further find that adverse selection result in "Third party" is driven by only those who are well informed of their own coverage. For "Theft" issues of small sample size prevents precise identification of the effects.

The findings are not coherent with the theory that knowledge is acquired only through claiming, or more generally after the insurance is purchased. This is because we find adverse selection is stronger among informed individuals even when our measure of riskiness is history of accidents prior to the contract or unclaimed accident during the contract.

One potential channel through which informed individuals make "better" choices is that they are also more informed about their own riskiness. This is not consistent with the findings: Table VI.1 shows that the interaction of adverse selection, knowledge and age - serving as a first proxy for knowing own risk - is positive and significant. This indicates that information has smaller effect over adverse selection for older insurees, who are more likely to know more about their riskiness. Similar results hold when I replace age with driving experience (Table VI.3), and become insignificant when replacing age with average distance driven per day (Table VI.4). These results still hold if I exclude extreme values of age, driving experience or average distance driven per day.

These results are potentially consistent with two alternative theories. Firstly, in case effort to understand coverage is costly, we would expect individuals who profit most from understanding coverage (and from making better choices) to acquire more information. In this case, we should see that individuals with characteristics that predict higher value of information about coverage choice - for instance, different levels of income, risk aversion, car valuation - would

[18] On the regression of measures of risk on knowledge dummy and selection dummy, the coefficient on knowledge dummy is significantly positive, while the coefficient on selection dummy is negative and significant. This suggests that there is adverse selection and informed individuals are riskier, however since informed individuals seem to choose less according to their riskiness than uninformed individuals, the interaction term becomes positive.

self-select better based on information. Secondly, information can be exogenous, in the sense that some individuals have better access or more capacity to understand insurance coverages.

Tables VI.6- VI.11 show preliminary results suggesting this is the case. In particular, individuals who respond more to information, in terms of selecting coverage coherent with their riskiness, have: 1) higher income, higher level of education and higher car valuation, and 2) bought insurance from sellers whom they know personally or through friends.

VI Conclusion

This paper intends to understand how adverse selection responds to consumers' information about insurance products. We consider this question to be particularly relevant in environments in which insurance firms face difficulties in using risk contingent pricing to control for adverse selection and moral hazard, and might thus be tempted to use consumer misinformation as a way of avoiding these informational problems.

I find that for third party risk, consumers who are most informed about their insurance coverage are the ones who are most likely to select coverage based on their riskiness. I argue this is consistent with two alternative theories: first, consumers who most profit from obtaining information in terms of their utility from insurance could be the ones with most incentive to acquire information. As a consequence, these individuals are also the ones for whom I most detect adverse selection. Secondly, different consumers might have different capacities or differential access to information for exogenous reasons. Based on this exogenous source of information, they end up with more "appropriate" coverage choice for their riskiness.

In future work, I plan to have a model consistent with the theories outlined above to qualitatively indicate the contexts in which these findings should be most relevant to policy. Also, such a model will be helpful in stating the trade-offs between consumer information and welfare versus adverse selection. This would help us in evaluating qualitatively whether 1) the firm has incentives to hide information from consumers in this environment and 2) whether policies advising consumers on insurance choices would be welfare improving.

References

[1] Abbring, J. H. et al. (2003). Moral Hazard and Dynamic Insurance Data. Journal of the European Economic Association, MIT Press, vol. 1(4), 767-820.

[2] Abbring, J. H. et al. (2003). Moral Hazard and Dynamic Insurance Data. Journal of the European Economic Association, MIT Press, vol. 1(4), 767-820.

[3] Akerlof, G.A. (1970). The brandet for Lemons: quality uncertainty and the brandet mechanism. Quarterly Journal of Economics, 84, 488500.

[4] Besley, T. (1995). Non-brandet Institutions for Credit and Risk Sharing in Low-Income Countries. Journal of Economic Perspectives, 9, 115127.

[5] Carter, M. R., Barrett, C. B. (2006). The Economics of Poverty Traps and Persistent Poverty: An Asset Based Approach. Journal of Development Studies, 42, 178199.

[6] Chiappori, P.-A., Salanié, B. (1997). Empirical Contract theory: The Case of Insurance Data. European Economic Review, Elsevier, vol. 41(3-5), 943-950.

[7] Chiappori, P.-A., Heckman, J. J. (1999). Testing for Moral Hazard Using Dynamic Data. Manuscript, Dept. of Economics, Univ. of Chicago.

[8] Chiappori, P.-A., Salanié, B. (2000). Testing for Asymmetric Information in Insurance brandets. Journal of Political Economy, University of Chicago Press, vol. 108(1), 56-78.

[9] Chiappori, P.-A. et al. (2006). Asymmetric Information in Insurance: General Testable Implications. RAND Journal of Economics, RAND Corporation, vol. 37(4), 783-798.

[10] Chiappori, P.-A., Salanié, B. (2008). Modeling Competition and brandet Equilibrium in Insurance: Empirical Issues. American Economic Review, American Economic Association, vol. 98(2), 146-50.

[11] Cohen, A. (2005). Asymmetric Information and Learning: Evidence from the Automobile Insurance brandet. Review of Economics and Statistics, Vol. 87 (2), 197-207.

[12] Cohen, A., Siegelman, P., (2010). Testing for Adverse Selection in Insurance brandets. Journal of Risk and Insurance, 77, 39-84.

[13] Cole, S. et al. (2013). Barriers to Household Risk Management: Evidence from India. American Economic Journal: Applied Economics, 5(1), 104-35.

[14] Cutler, D. M. et al. (2008). Preference Heterogeneity and Insurance brandets: Explaining a Puzzle of Insurance. American Economic Review, American Economic Association, vol. 98(2), 157-62.

[15] Dahlby, B.G. (1983). Adverse Selection and Statistical Discrimination, an Analysis of Canadian Automobile Insurance. Journal of Public Economics, 20, 121-130.

[16] Dahlby, B.G. (1992). Testing for Asymmetric Information in Canadian Automobile insurance. Contributions to Insurance Economics, edited by Georges Dionne, Boston: Kluwer

[17] De Meza, D. & Webb, D. C. (2001). Advantageous Selection in Insurance brandets. RAND Journal of Economics, The RAND Corporation, vol. 32(2), 249-62.

[18] Dionne, G., Doherty, N., Fombaron, N., (2000). Adverse Selection in Insurance brandets. In: G. Dionne, editor, Handbook of Insurance, Chapter 7, 185-244. Kluwer Academic Publishers.

[19] Dionne, G., Gourieroux C., Vanasse, C. (2001). Testing for Evidence of Adverse Selection in the Automobile Insurance brandet: A Comment. The Journal of Political Economy 109, 444-453.

[20] Dionne, G., Michaud, P.-C., Dahchour, M., 2004. Separating Moral Hazard from Adverse Selection in Automobile Insurance: Longitudinal Evidence from France. Cahiers de recherche 0420, CIRPEE.

[21] Einav, L. et al. (2010). Beyond Testing: Empirical Models of Insurance brandets. Annual Review of Economics, Annual Reviews, vol. 2(1), 311-336.

[22] Finkelstein, A., McGarry, K. (2006). Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance brandet. American Economic Review, American Economic Association, vol. 96(4), 938-958.

[23] Finkelstein, A. et al. (2012). The Oregon Health Insurance Experiment: Evidence from the First Year. The Quarterly Journal of Economics, Oxford University Press, vol. 127(3), 1057-1106.

[24] Gunnsteinsson, S. (2014). Identifying Information Asymmetries in Insurance: Theory and Evidence on Crop Insurance in the Philippines. Working paper.

[25] Karlan, D., Zinman, J. (2009). Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. Econometrica, Econometric Society, vol. 77(6), 1993-2008.

[26] Kim, H. et al. (2009). Evidence of Asymmetric Information in the Automobile Insurance brandet: Dichotomous Versus Multinomial Measurement of Insurance Coverage. The Journal of Risk and Insurance, 76, 343-366.

[27] Kim, Y. S. (2013). Consumer Search Frictions, Competition and Adverse Selection in Health Insurance brandets: Evidence from Medigap. Working paper.

[28] Handel, B. R. (2011). Adverse Selection and Switching Costs in Health Insurance brandets: When Nudging Hurts. NBER Working Papers 17459, National Bureau of Economic Research, Inc.

[29] Handel, B. R. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. American Economic Review, 103 (7), 2643-82.

[30] Puelz, R., Snow, A. (1994). Evidence on Adverse Selection: Equilibrium Signalling and Cross-Subsidization in the Insurance brandet. Journal of Political Economy, vol. 102 (2), 236-257.

[31] Richaudeau, D. (1999). Automobile Insurance Contracts and Risk of Accident: an Empirical Test Using French Individual Data. The Geneva Papers on Risk and Insurance - Theory 24 (1), 97-114.

[32] Rothschild, M., Stiglitz, J.E. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. The Quarterly Journal of Economics, MIT Press, vol. 90(4), 630-49.

[33] Spinnewijn, J. (2013), Insurance and Perceptions: How to Screen Optimists and Pessimists. The Economic Journal, 123: 606633

[34] Starc, A. (2014). Insurer Pricing and Consumer Welfare: Evidence from Medigap. The RAND Journal of Economics, 45 (1), 198220.

Appendices

I Additional notes on the institutional background

A Bonus structure: agents vs managers

Around 100 agents were registered with Ulaanbaatar branches as of July 2013, however, only 10-15 turned out to be active sellers. This is due to a distortive bonus mechanism at the firm, with agents receiving a bonus of around 15% of generated income, while managers receive around 1.8%. Furthermore, with differentiated bonus structure, managers often write contracts in the name of agents', whom they bring in, and collect their bonuses. Banks and brokers can also do so[19]. Due to income-based bonus system, lack of underwriting skills and limited monitoring, sellers do not underwrite or evaluate risk, that is their incentives are misaligned with those of the firm.

B "Bonus malus" within the contract

Uncommon type of "bonus malus" exists at MD, with the current coverage reducing with each additional claim, but only in the duration of the current contract. Claims in the current contract do not affect next period pricing. A brief explanation is the following. An insuree can claim any amount up to the vehicle valuation. The firm promises to pay out an amount equivalent to:

$$\text{Loss val} * (1-c) * \min\left(1, \frac{\text{Initial car val} - \sum \text{Past claims}}{\text{Car val at the time of accident}}\right) \tag{5}$$

where c is the co-insurance rate. Hence with each claim the coverage within the same contract decreases, with the insuree being allowed to "top-up" insurance.

C Claims procedure

Generally, an insuree may need to provide different sets of materials in the claim application, depending on the type of risk incurred. The insurance company will only consider claim applications if no road traffic laws are violated by the insuree and the insuree must inform the firm hotline of the accident

[19]This prevents researchers from identifying the actual seller and the volume of sales through each channel. Only after the project has started it became clear that most contracts were sold through banks and brokers.

while remaining at the incident location[20]. The hotline team arrives at the location to inspect the situation, make notes and inform of the set of documents the insuree needs to provide for a claim application. Police reports are often requested to determine the faulty party, so that the insurance company can rebate the claim from the faulty third party (subrogation). The insuree may then choose the damage evaluator, who reports both the loss and vehicle valuation at the time of the incident. These enter in the payout equation (5). Survey respondents were distrustful of and dissatisfied with the loss and the car valuations, suspecting firms might be colluding with damage evaluators. From discussions with interviewers, it was clear that many insurees also thought that they could only claim once during the coverage period, so they delayed claiming until loss size was justifiable.

D Seller's manual with randomisation in place

All sellers underwent training and were given a manual on how to finalise insurance sales with the randomisation in place. Few points from the manual worth mentioning are:

1. Sellers should explain carefully what each coverage insures against.
2. Once the contract is signed and payment is finalised, both the seller and the buyer send a text message to a server using their mobile phones (with some additional information in the messages) and if the process is accepted by the server, they both receive messages with the randomisation outcome.
3. The seller should note the randomisation outcome on the contract and have the insuree sign in the box next to it.
4. If an insuree is upgraded to a higher coverage, the seller should explain carefully what this implies.
5. If the message is sent before the contract starts or before a contract is finalised, then the employee faces a punishment equivalent to a month's minimum wage per message.

[20]This is in line with police requirements to not move from the location of the accident due lack of/low quality road traffic cameras, even on the main roads.

II Balance tables for the randomisation

	No upgrade	Upgrade	std.diff	z	
Male	0.65	0.70	0.12	1.03	
Age	39.47	39.33	-0.01	-0.10	
factor(Resid)Bgd	0.15	0.16	0.03	0.30	
factor(Resid)Bzd	0.22	0.19	-0.08	-0.68	
factor(Resid)Chd	0.10	0.13	0.08	0.70	
factor(Resid)Hud	0.15	0.15	-0.01	-0.09	
factor(Resid)Other	0.05	0.06	0.02	0.20	
factor(Resid)Sbd	0.12	0.11	-0.01	-0.13	
factor(Resid)Shd	0.20	0.20	-0.01	-0.08	
Car age	11.56	11.64	0.02	0.18	
Car valuation	11552985.07	13821954.17	0.13	1.19	
factor(Brand_contin)Asia	0.95	0.95	0.02	0.22	
factor(Brand_contin)Europe	0.03	0.04	0.06	0.56	
factor(Brand_contin)Unknown	0.01	-0.00	-0.19	-1.70	.
factor(Brand_contin)US	0.01	0.01	-0.03	-0.26	
Lightcolour	0.63	0.71	0.17	1.51	
worker_male	0.28	0.25	-0.08	-0.68	
worker_age	35.08	34.01	-0.10	-0.88	
worker_exper	3.30	2.98	-0.07	-0.62	
worker_agent	0.19	0.09	-0.28	-2.43	*
Days insured	148.79	175.05	0.27	2.38	*

Table II.1: Balance tables: those who stayed in product 1 cover against "Vehicle accident" *versus* those who were upgraded from product 1 to product 2 cover against "Vehicle accident" and "Third party". The first column shows the pre-randomisation variables. I have included seller-specific characteristics and days insured to see if certain sellers sold more upgraded contracts than others. Days insured is the length of time for which the contract was valid until July 2014.

	No upgrade	Upgrade	std.diff	z	
Male	0.61	0.64	0.08	0.86	
Age	39.02	38.62	-0.03	-0.39	
factor(Resid)Bgd	0.25	0.22	-0.06	-0.68	
factor(Resid)Bzd	0.16	0.19	0.08	0.95	
factor(Resid)Chd	0.11	0.13	0.08	0.90	
factor(Resid)Hud	0.15	0.13	-0.05	-0.54	
factor(Resid)Other	0.02	0.03	0.10	1.11	
factor(Resid)Sbd	0.12	0.12	-0.01	-0.09	
factor(Resid)Shd	0.20	0.17	-0.07	-0.76	
Car age	10.65	10.67	0.00	0.05	
Car valuation	13995352.80	13638082.90	-0.02	-0.25	
factor(Brand_contin)Asia	0.95	0.96	0.07	0.81	
factor(Brand_contin)Europe	0.04	0.02	-0.11	-1.29	
factor(Brand_contin)US	0.01	0.02	0.05	0.62	
Lightcolour	0.74	0.74	-0.00	-0.04	
worker_male	0.22	0.24	0.05	0.53	
worker_age	34.70	33.04	-0.16	-1.79	.
worker_exper	4.02	4.42	0.08	0.89	
worker_agent	0.25	0.12	-0.30	-3.43	***
Days insured	164.80	177.07	0.12	1.36	

Table II.2: Balance tables: those who stayed in product 2 (10% co-insurance rate) cover against "Vehicle accident" and "Third party" *versus* those who were upgraded from product 2 to product 3 (10% co-insurance rate) cover against "Vehicle accident", "Third party" and "Theft". The first column shows the pre-randomisation variables. I have included seller-specific characteristics and days insured to see if certain sellers sold more upgraded contracts than others. Days insured is the length of time for which the contract was valid until July 2014.

	No upgrade	Upgrade	std.diff	z	
Male	0.61	0.65	0.09	1.08	
Age	39.02	38.85	-0.01	-0.17	
factor(Resid)Bgd	0.25	0.22	-0.07	-0.82	
factor(Resid)Bzd	0.16	0.17	0.04	0.45	
factor(Resid)Chd	0.11	0.13	0.07	0.79	
factor(Resid)Hud	0.15	0.14	-0.03	-0.34	
factor(Resid)Other	0.02	0.03	0.11	1.33	
factor(Resid)Sbd	0.12	0.10	-0.08	-0.91	
factor(Resid)Shd	0.20	0.21	0.03	0.34	
Car age	10.65	10.84	0.05	0.60	
Car valuation	13995352.80	13726779.93	-0.02	-0.19	
factor(Brand_contin)Asia	0.95	0.97	0.11	1.26	
factor(Brand_contin)Europe	0.04	0.03	-0.06	-0.76	
factor(Brand_contin)US	0.01	0.00	-0.12	-1.42	
Lightcolour	0.74	0.71	-0.06	-0.75	
worker_male	0.22	0.30	0.19	2.23	*
worker_age	34.70	33.42	-0.12	-1.42	
worker_exper	4.02	4.03	0.00	0.02	
worker_agent	0.25	0.12	-0.32	-3.76	***
Days insured	164.80	195.55	0.30	3.48	***

Table II.3: Balance tables: those who stayed in product 2 (10% co-insurance rate) cover against "Vehicle accident" and "Third party" *versus* those who were upgraded from product 2 (10% co-insurance rate) to product 2 (no co-insurance rate) cover against "Vehicle accident" and "Third party". The first column shows the pre-randomisation variables. I have included seller-specific characteristics and days insured to see if certain sellers sold more upgraded contracts than others. Days insured is the length of time for which the contract was valid until July 2014.

III Survey: balance tables

	Group A	Group C	std.diff	z	
Male	0.61	0.65	0.07	1.10	
Age	40.37	37.62	-0.22	-3.63	***
Upgrade	0.44	0.43	-0.03	-0.46	
Claim freq	0.29	0.28	-0.03	-0.41	
Claim prob	0.21	0.21	0.00	0.08	
Claim size	153348.33	141440.85	-0.01	-0.22	
Collision freq	0.17	0.20	0.05	0.76	
Collision prob	0.14	0.15	0.04	0.65	
Collision claim	130953.95	121996.18	-0.01	-0.17	
Third party freq	0.11	0.07	-0.12	-2.01	*
Third party prob	0.10	0.07	-0.10	-1.66	.
Third party claim	21568.19	17988.93	-0.04	-0.71	
Theft freq	0.01	0.01	0.01	0.17	
Theft prob	0.01	0.01	0.01	0.17	
Theft claim	826.19	1455.73	0.03	0.54	
Car age	10.97	10.90	-0.02	-0.33	
Car valuation	12747790.16	14996322.58	0.14	2.23	*
Days insured	173.92	178.69	0.05	0.75	
factor(Resid)Bgd	0.24	0.16	-0.21	-3.34	***
factor(Resid)Bzd	0.18	0.18	-0.02	-0.32	
factor(Resid)Chd	0.13	0.11	-0.05	-0.76	
factor(Resid)Hud	0.12	0.18	0.15	2.39	*
factor(Resid)Other	0.00	0.09	0.46	7.25	***
factor(Resid)Sbd	0.12	0.12	-0.00	-0.04	
factor(Resid)Shd	0.20	0.17	-0.09	-1.47	
factor(Brand_contin)Asia	0.96	0.95	-0.05	-0.78	
factor(Brand_contin)Europe	0.03	0.04	0.04	0.59	
factor(Brand_contin)Unknown	0.00	0.00	0.07	1.07	
factor(Brand_contin)US	0.01	0.01	0.01	0.21	
Lightcolour	0.72	0.72	0.00	0.01	
worker_male	0.27	0.26	-0.02	-0.39	
worker_age	33.54	34.44	0.09	1.40	
worker_exper	4.03	4.08	0.01	0.15	
worker_agent	0.15	0.15	-0.01	-0.09	
Seller_friend	0.27	0.26	-0.03	-0.56	

Table III.1: Balance tables comparing those who participated in the survey (group A) *versus* potential survey respondents, including individuals who we were not able to reach by phone or those agreed to be interviewed outside the survey timeline (group B).

	Group A	Group B	std.diff	z	
Male	0.61	0.61	-0.00	-0.05	
Age	40.37	38.12	-0.19	-2.73	**
Upgrade	0.44	0.39	-0.11	-1.64	
Claim freq	0.29	0.32	0.05	0.68	
Claim prob	0.21	0.22	0.02	0.34	
Claim size	153348.33	117936.54	-0.04	-0.62	
Collision freq	0.17	0.21	0.08	1.15	
Collision prob	0.14	0.16	0.06	0.82	
Collision claim	130953.95	92193.27	-0.05	-0.69	
Collision freq	0.11	0.11	-0.02	-0.22	
Third party prob	0.10	0.09	-0.01	-0.09	
Third party claim	21568.19	21832.26	0.00	0.05	
Theft freq	0.01	0.00	-0.03	-0.48	
Theft prob	0.01	0.00	-0.03	-0.48	
Theft claim	826.19	1211.31	0.02	0.34	
Car age	10.97	10.58	-0.10	-1.46	
Car valuation	12747790.16	15996468.50	0.20	2.88	**
Days insured	173.92	171.14	-0.03	-0.38	
factor(Resid)Bgd	0.24	0.23	-0.02	-0.22	
factor(Resid)Bzd	0.18	0.20	0.04	0.52	
factor(Resid)Chd	0.13	0.11	-0.06	-0.88	
factor(Resid)Hud	0.12	0.14	0.05	0.68	
factor(Resid)Sbd	0.12	0.12	0.01	0.12	
factor(Resid)Shd	0.20	0.20	-0.02	-0.23	
factor(Brand_contin)Asia	0.96	0.96	-0.01	-0.17	
factor(Brand_contin)Europe	0.03	0.03	-0.01	-0.09	
factor(Brand_contin)Unknown	0.00	0.00	0.09	1.32	
factor(Brand_contin)US	0.01	0.01	0.00	0.06	
Lightcolour	0.72	0.68	-0.09	-1.34	
worker_male	0.27	0.23	-0.11	-1.58	
worker_age	33.54	34.58	0.10	1.43	
worker_exper	4.03	4.14	0.02	0.28	
worker_agent	0.15	0.17	0.06	0.84	
Seller_friend	0.27	0.28	0.02	0.35	

Table III.2: Balance tables comparing those who participated in the survey (group A) *versus* those who refused to participate (group B).

	Group A	Group B, C	std.diff	z	
Male	0.61	0.63	0.04	0.72	
Age	40.37	37.82	-0.22	-3.93	***
Upgrade	0.44	0.41	-0.06	-1.14	
Claim freq	0.29	0.29	0.00	0.09	
Claim prob	0.21	0.21	0.01	0.22	
Claim size	153348.33	132113.29	-0.03	-0.50	
Collision freq	0.17	0.20	0.06	1.08	
Collision prob	0.14	0.16	0.05	0.85	
Collision claim	130953.95	110169.05	-0.03	-0.49	
Third party freq	0.11	0.09	-0.08	-1.45	
Third party prob	0.10	0.08	-0.06	-1.16	
Third party claim	21568.19	19514.14	-0.03	-0.46	
Theft freq	0.01	0.00	-0.01	-0.11	
Theft prob	0.01	0.00	-0.01	-0.11	
Theft claim	826.19	1358.74	0.03	0.49	
Car age	10.97	10.77	-0.05	-0.95	
Car valuation	12747790.16	15393225.15	0.15	2.73	**
Days insured	173.92	175.69	0.02	0.31	
factor(Resid)Bgd	0.24	0.19	-0.13	-2.35	*
factor(Resid)Bzd	0.18	0.19	0.00	0.05	
factor(Resid)Chd	0.13	0.11	-0.05	-0.97	
factor(Resid)Hud	0.12	0.16	0.11	1.96	.
factor(Resid)Other	0.00	0.05	0.31	5.60	***
factor(Resid)Sbd	0.12	0.12	0.00	0.04	
factor(Resid)Shd	0.20	0.18	-0.06	-1.11	
factor(Brand_contin)Asia	0.96	0.95	-0.03	-0.61	
factor(Brand_contin)Europe	0.03	0.04	0.02	0.36	
factor(Brand_contin)Unknown	0.00	0.00	0.06	1.18	
factor(Brand_contin)US	0.01	0.01	0.01	0.18	
Lightcolour	0.72	0.71	-0.04	-0.67	
worker_male	0.27	0.25	-0.06	-1.06	
worker_age	33.54	34.50	0.09	1.66	.
worker_exper	4.03	4.11	0.01	0.24	
worker_agent	0.15	0.16	0.02	0.37	
Seller_friend	0.27	0.27	-0.01	-0.20	

Table III.3: Balance tables comparing those who participated in the survey (group A) *versus* all other insurees (group B, C).

	Group A	Group B, C	std.diff	z	
Male	0.62	0.63	0.03	0.56	
Age	40.40	37.59	-0.24	-4.31	***
Upgrade	0.44	0.42	-0.05	-0.97	
Claim freq	0.29	0.30	0.01	0.17	
Claim prob	0.21	0.21	0.02	0.29	
Claim size	153889.24	134355.89	-0.03	-0.45	
Collision freq	0.17	0.21	0.07	1.20	
Collision prob	0.14	0.16	0.05	0.99	
Collision claim	131415.87	112629.40	-0.02	-0.44	
Third party freq	0.11	0.09	-0.08	-1.49	
Third party prob	0.10	0.08	-0.07	-1.23	
Third party claim	21644.27	19242.12	-0.03	-0.54	
Theft freq	0.01	0.00	-0.00	-0.08	
Theft prob	0.01	0.00	-0.00	-0.08	
Theft claim	829.10	1389.08	0.03	0.51	
Car age	11.00	10.98	-0.01	-0.13	
Car valuation	12404748.85	12941602.14	0.05	0.86	
Days insured	174.08	176.25	0.02	0.38	
factor(Resid)Bgd	0.24	0.19	-0.13	-2.34	*
factor(Resid)Bzd	0.19	0.19	0.01	0.10	
factor(Resid)Chd	0.13	0.12	-0.05	-0.84	
factor(Resid)Hud	0.13	0.15	0.08	1.50	
factor(Resid)Other	0.00	0.05	0.31	5.65	***
factor(Resid)Sbd	0.12	0.12	0.01	0.12	
factor(Resid)Shd	0.20	0.18	-0.05	-0.95	
factor(Brand_contin)Asia	0.96	0.96	-0.01	-0.26	
factor(Brand_contin)Europe	0.03	0.03	0.00	0.05	
factor(Brand_contin)Unknown	0.00	0.00	0.07	1.19	
factor(Brand_contin)US	0.01	0.01	-0.00	-0.03	
Lightcolour	0.73	0.72	-0.02	-0.39	
worker_male	0.27	0.25	-0.06	-1.08	
worker_age	33.56	34.53	0.09	1.67	.
worker_exper	4.04	4.08	0.01	0.13	
worker_agent	0.15	0.16	0.02	0.39	
Seller_friend	0.27	0.26	-0.02	-0.38	

Table III.4: Balance tables comparing those who participated in the survey (group A) *versus* all other insurees (group B, C), excluding top 1% highest values for car valuation.

IV Adverse selection and moral hazard

A Adverse selection in “Third party”

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.045**	-0.049**	-0.041**	-0.043**	-0.503**	-0.515**
	(0.021)	(0.020)	(0.019)	(0.019)	(0.231)	(0.245)
log(Days insured)	0.087***	0.095***	0.067***	0.072***	0.572***	0.597***
	(0.021)	(0.022)	(0.017)	(0.018)	(0.095)	(0.094)
Male		0.044**		0.034*		0.450**
		(0.019)		(0.017)		(0.222)
Age		-0.003***		-0.002**		-0.024***
		(0.001)		(0.001)		(0.008)
log(Carval)		0.004		0.010		0.140
		(0.020)		(0.018)		(0.179)
Other controls	NO	YES	NO	YES	NO	YES
N	988	977	988	977	988	977
R^2					0.025	0.049
Log Likelihood	-300.880	-283.202	-272.195	-255.471		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.1: Standard errors are in parentheses. The dependent variables are third party accident frequency, probability and claim size from the administrative claims data. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice, controlling for *log*(*Days insured*) only. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.055***	-0.058***	-0.050***	-0.052***	-0.629***	-0.635***
	(0.020)	(0.020)	(0.018)	(0.018)	(0.223)	(0.237)
log(Days insured)	0.090***	0.098***	0.069***	0.074***	0.581***	0.603***
	(0.021)	(0.023)	(0.018)	(0.018)	(0.098)	(0.096)
Male		0.049**		0.038**		0.517**
		(0.019)		(0.017)		(0.220)
Age		-0.002**		-0.002**		-0.022***
		(0.001)		(0.001)		(0.008)
log(Carval)		0.004		0.010		0.155
		(0.021)		(0.018)		(0.180)
Other controls	NO	YES	NO	YES	NO	YES
N	965	954	965	954	965	954
R^2					0.027	0.051
Log Likelihood	-290.834	-273.448	-262.181	-245.759		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.2: Standard errors are in parentheses. The dependent variables are third party accident frequency, probability and claim size from the administrative claims data. For this regression, I exclude 23 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice, controlling for *log*(*Days insured*) only. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

B Adverse selection in "Theft"

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.015*	0.017	0.015*	0.016*	0.145*	0.134
	(0.008)	(0.010)	(0.008)	(0.009)	(0.087)	(0.142)
log(Days insured)	0.057	0.070	0.056	0.068	0.145*	0.154*
	(0.057)	(0.065)	(0.056)	(0.063)	(0.084)	(0.088)
Male		0.00005		0.0004		0.008
		(0.017)		(0.016)		(0.213)
Age		-0.0004		-0.0004		0.001
		(0.001)		(0.001)		(0.003)
log(Carval)		0.013		0.013		0.019
		(0.025)		(0.024)		(0.076)
Other controls	NO	YES	NO	YES	NO	YES
N	238	233	238	233	238	233
R^2					0.012	0.018
Log Likelihood	-13.073	-11.648	-13.010	-11.510		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.3: Standard errors are in parentheses. The dependent variables are theft accident frequency, probability and claim size from the administrative claims data. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice, controlling for $log(Days\ insured)$ only. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.016*	0.018	0.016*	0.017*	0.154*	0.141
	(0.009)	(0.011)	(0.009)	(0.010)	(0.092)	(0.146)
log(Days insured)	0.060	0.072	0.059	0.071	0.155*	0.162*
	(0.059)	(0.067)	(0.058)	(0.065)	(0.090)	(0.093)
Male		0.0001		0.0004		0.009
		(0.018)		(0.016)		(0.222)
Age		-0.0004		-0.0004		0.001
		(0.001)		(0.001)		(0.003)
log(Carval)		0.014		0.014		0.020
		(0.027)		(0.026)		(0.078)
Other controls	NO	YES	NO	YES	NO	YES
N	226	221	226	221	226	221
R^2					0.013	0.018
Log Likelihood	-12.854	-11.546	-12.786	-11.407		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.4: Standard errors are in parentheses. The dependent variables are theft accident frequency, probability and claim size from the administrative claims data. For this regression, I exclude 12 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice, controlling for $log(Days\ insured)$ only. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.102	-0.118	-0.116	-0.161	0.278	0.087
	(0.109)	(0.123)	(0.095)	(0.107)	(0.200)	(0.138)
log(Days insured)	0.092	0.088	0.076	0.078	0.240	0.292
	(0.063)	(0.064)	(0.053)	(0.053)	(0.170)	(0.197)
Male		-0.002		-0.004		0.392
		(0.078)		(0.064)		(0.287)
Age		-0.006		-0.005		-0.003
		(0.004)		(0.004)		(0.006)
log(Carval)		-0.112		-0.121		-0.089
		(0.094)		(0.084)		(0.159)
Other controls	NO	YES	NO	YES	NO	YES
N	97	97	97	97	97	97
R^2					0.021	0.068
Log Likelihood	-35.490	-30.937	-31.746	-26.068		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.5: Standard errors are in parentheses. The dependent variables are theft accident frequency, probability and loss size **during the contract** from the survey data. These include both claimed and unclaimed accidents. For this regression, I exclude 5 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice, controlling for $log(Days\ insured)$ only. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.211	-0.361	-0.097	-0.144	-1.004	-1.154
	(0.148)	(0.256)	(0.103)	(0.122)	(1.240)	(1.191)
log(Days insured)	-0.027	-0.017	-0.046	-0.043	-0.319	-0.376
	(0.050)	(0.056)	(0.041)	(0.041)	(0.526)	(0.582)
Male		0.201		0.115		1.124
		(0.138)		(0.079)		(0.913)
Age		-0.004		-0.001		-0.001
		(0.005)		(0.003)		(0.032)
log(Carval)		0.060		0.068		0.851
		(0.100)		(0.075)		(0.867)
Other controls	NO	YES	NO	YES	NO	YES
N	97	97	97	97	97	97
R^2					0.014	0.084
Log Likelihood	-58.022	-52.025	-42.128	-39.586		

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.6: Standard errors are in parentheses. The dependent variables are history of theft accident frequency, probability and loss size **prior to the contract** from the survey data. These include both claimed and unclaimed accidents. For this regression, I exclude 5 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice, controlling for $log(Days\ insured)$ only. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.314*	-0.465*	-0.229*	-0.295**	-0.737	-1.086
	(0.184)	(0.258)	(0.121)	(0.129)	(1.250)	(1.200)
Male		0.162		0.085		1.533
		(0.143)		(0.093)		(0.945)
Age		-0.010		-0.006		-0.004
		(0.006)		(0.004)		(0.034)
log(Carval)		-0.012		-0.014		0.738
		(0.131)		(0.091)		(0.832)
Other controls	NO	YES	NO	YES	NO	YES
N	97	97	97	97	97	97
R^2					0.004	0.068
Log Likelihood	-75.656	-70.853	-54.410	-51.991		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.7: Standard errors are in parentheses. The dependent variables are history of theft accident frequency, probability and loss size prior to the interview date from the survey data. These include both claimed and unclaimed accidents. For this regression, I exclude 5 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice without any additional controls. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.314*	-0.459*	-0.229*	-0.302**	-0.737	-1.161
	(0.184)	(0.236)	(0.121)	(0.127)	(1.250)	(1.269)
Male		0.118		0.093		1.114
		(0.128)		(0.090)		(0.847)
Age		-0.011*		-0.006		-0.005
		(0.006)		(0.004)		(0.032)
Other controls	NO	NO	NO	NO	NO	NO
N	97	97	97	97	97	97
R^2					0.004	0.018
Log Likelihood	-75.656	-73.171	-54.410	-52.932		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.8: Standard errors are in parentheses. The dependent variables are history of theft accident frequency, probability and loss size prior to the interview date from the survey data. These include both claimed and unclaimed accidents. For this regression, I exclude 5 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice without any additional controls. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for gender and age of the insuree.

C Adverse selection in coinsurance rate

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.015	-0.029	0.006	-0.003	0.038	-0.123
	(0.060)	(0.063)	(0.043)	(0.044)	(0.540)	(0.550)
log(Days insured)	0.320***	0.345***	0.181***	0.189***	1.604***	1.694***
	(0.068)	(0.072)	(0.044)	(0.045)	(0.229)	(0.245)
Male		0.030		0.005		-0.047
		(0.064)		(0.045)		(0.564)
Age		-0.005		-0.002		-0.029
		(0.003)		(0.002)		(0.023)
log(Carval)		-0.124*		-0.042		-0.478
		(0.066)		(0.044)		(0.462)
Other controls	NO	YES	NO	YES	NO	YES
N	374	370	374	370	374	370
R^2					0.064	0.095
Log Likelihood	-264.177	-254.882	-189.477	-182.621		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.9: Standard errors are in parentheses. The dependent variables are "Vehicle accident" and "Third party" accident frequency, probability and claim size from the administrative claims data. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.011	-0.028	0.007	-0.004	0.046	-0.139
	(0.061)	(0.064)	(0.043)	(0.044)	(0.547)	(0.558)
log(Days insured)	0.346***	0.376***	0.197***	0.208***	1.711***	1.828***
	(0.073)	(0.077)	(0.047)	(0.049)	(0.238)	(0.257)
Male		0.013		-0.011		-0.198
		(0.066)		(0.047)		(0.580)
Age		-0.003		-0.001		-0.017
		(0.003)		(0.002)		(0.024)
log(Carval)		-0.131*		-0.045		-0.501
		(0.067)		(0.045)		(0.468)
Other controls	NO	YES	NO	YES	NO	YES
N	362	358	362	358	362	358
R^2					0.070	0.097
Log Likelihood	-257.191	-248.803	-182.745	-176.655		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.10: Standard errors are in parentheses. The dependent variables are "Vehicle accident" and "Third party" accident frequency, probability and claim size from the administrative claims data. For this regression, I exclude 12 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of riskiness on coverage choice controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

D Moral hazard in "Third party"

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	-0.015	0.002	-0.021	-0.002	1.007**	1.061**
	(0.094)	(0.099)	(0.078)	(0.074)	(0.395)	(0.417)
log(Days insured)	0.050	0.085	0.038	0.055	0.287	0.391
	(0.056)	(0.061)	(0.045)	(0.045)	(0.226)	(0.248)
Male		-0.402**		-0.276***		-1.174
		(0.159)		(0.088)		(0.718)
Age		0.005		0.004		0.008
		(0.004)		(0.003)		(0.021)
log(Carval)		0.028		0.013		-0.207
		(0.081)		(0.059)		(0.272)
Other controls	NO	YES	NO	YES	NO	YES
N	118	118	118	118	118	118
R^2					0.046	0.111
Log Likelihood	-71.336	-63.025	-61.829	-52.479		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.11: Standard errors are in parentheses. The dependent variables are third party accident frequency, probability and loss size **during the contract** from the survey data. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	-0.058	-0.043	-0.067	-0.052	0.708**	0.761**
	(0.093)	(0.099)	(0.077)	(0.076)	(0.345)	(0.365)
log(Days insured)	0.056	0.098	0.044	0.066	0.191	0.253
	(0.057)	(0.064)	(0.047)	(0.049)	(0.198)	(0.220)
Male		-0.349**		-0.204**		-0.580
		(0.165)		(0.090)		(0.650)
Age		0.005		0.004		0.011
		(0.004)		(0.003)		(0.022)
log(Carval)		0.045		0.031		-0.004
		(0.074)		(0.057)		(0.134)
Other controls	NO	YES	NO	YES	NO	YES
N	112	112	112	112	112	112
R^2					0.033	0.073
Log Likelihood	-63.888	-56.933	-54.698	-47.061		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.12: Standard errors are in parentheses. The dependent variables are third party accident frequency, probability and loss size **during the contract** from the survey data. For this regression, I exclude 6 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

E Moral hazard in "Theft"

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	0.053	0.056	0.047	0.048	0.288	0.260
	(0.039)	(0.042)	(0.035)	(0.035)	(0.201)	(0.181)
log(Days insured)	0.071**	0.076**	0.057*	0.060*	0.091	0.086
	(0.035)	(0.037)	(0.029)	(0.031)	(0.064)	(0.060)
Male		0.059*		0.046*		0.115
		(0.031)		(0.028)		(0.088)
Age		-0.003		-0.003		-0.002
		(0.002)		(0.002)		(0.002)
log(Carval)		0.003		-0.001		-0.012
		(0.046)		(0.039)		(0.075)
Other controls	NO	YES	NO	YES	NO	YES
N	255	254	255	254	255	254
R^2					0.023	0.037
Log Likelihood	-61.934	-55.698	-55.215	-49.498		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.13: Standard errors are in parentheses. The dependent variables are theft accident frequency, probability and loss size **during the contract** from the survey data. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	0.046	0.043	0.040	0.036	0.303	0.278
	(0.039)	(0.040)	(0.035)	(0.034)	(0.211)	(0.193)
log(Days insured)	0.065*	0.066*	0.051*	0.051*	0.091	0.087
	(0.034)	(0.035)	(0.029)	(0.029)	(0.064)	(0.061)
Male		0.051*		0.038		0.117
		(0.031)		(0.028)		(0.089)
Age		-0.003		-0.003		-0.002
		(0.002)		(0.002)		(0.002)
log(Carval)		-0.003		-0.006		-0.012
		(0.045)		(0.038)		(0.077)
Other controls	NO	YES	NO	YES	NO	YES
N	250	249	250	249	250	249
R^2					0.024	0.038
Log Likelihood	-59.773	-53.079	-53.086	-46.950		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.14: Standard errors are in parentheses. The dependent variables are theft accident frequency, probability and loss size **during the contract** from the survey data. For this regression, I exclude 5 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

F Moral hazard in coinsurance rate

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	-0.010	-0.030	0.005	-0.007	0.086	-0.093
	(0.045)	(0.045)	(0.033)	(0.032)	(0.431)	(0.439)
log(Days insured)	0.325***	0.341***	0.178***	0.182***	1.614***	1.619***
	(0.050)	(0.053)	(0.032)	(0.033)	(0.178)	(0.177)
Male		0.095**		0.047		0.626
		(0.044)		(0.032)		(0.412)
Age		-0.007***		-0.005***		-0.056***
		(0.002)		(0.002)		(0.016)
log(Carval)		-0.020		-0.012		-0.086
		(0.049)		(0.034)		(0.327)
Other controls	NO	YES	NO	YES	NO	YES
N	629	622	629	622	629	622
R^2					0.074	0.114
Log Likelihood	-410.314	-392.420	-299.526	-282.748		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.15: Standard errors are in parentheses. The dependent variables are "Vehicle accident" and "Third party" accident frequency, probability and claim size from the adminstrative claims data. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	-0.008	-0.030	0.005	-0.010	0.089	-0.119
	(0.046)	(0.046)	(0.034)	(0.033)	(0.440)	(0.452)
log(Days insured)	0.340***	0.357***	0.187***	0.192***	1.672***	1.679***
	(0.052)	(0.055)	(0.034)	(0.035)	(0.182)	(0.181)
Male		0.089**		0.041		0.551
		(0.045)		(0.032)		(0.419)
Age		-0.006***		-0.004***		-0.051***
		(0.002)		(0.002)		(0.016)
log(Carval)		-0.022		-0.013		-0.082
		(0.050)		(0.034)		(0.330)
Other controls	NO	YES	NO	YES	NO	YES
N	617	610	617	610	617	610
R^2					0.078	0.114
Log Likelihood	-403.138	-386.759	-292.707	-277.507		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.16: Standard errors are in parentheses. The dependent variables are "Vehicle accident" and "Third party" accident frequency, probability and claim size from the administrative claims data. For this regression, I exclude 12 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Upgrade	0.098	0.130	-0.015	0.001	0.219	0.334
	(0.112)	(0.116)	(0.063)	(0.062)	(0.640)	(0.644)
log(Days insured)	0.326***	0.377***	0.129***	0.140***	1.496***	1.481***
	(0.080)	(0.086)	(0.041)	(0.041)	(0.266)	(0.258)
Male		-0.007		-0.001		1.259**
		(0.115)		(0.062)		(0.582)
Age		-0.019***		-0.007***		-0.033
		(0.005)		(0.003)		(0.023)
log(Carval)		-0.290**		-0.135*		-0.900*
		(0.138)		(0.073)		(0.502)
Other controls	NO	YES	NO	YES	NO	YES
N	267	266	267	266	267	266
R^2					0.070	0.105
Log Likelihood	-318.581	-301.342	-177.893	-168.113		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table IV.17: Standard errors are in parentheses. The dependent variables are "Vehicle accident" and "Third party" accident frequency, probability and claim size **during the contract** from the survey data. For this regression, I exclude 7 contracts potentially biasing the randomisation. Row 1 of columns 1, 3 and 5 show estimates of effect of high coverage on riskiness controlling for $log(Days\ insured)$. Row 1 of columns 2, 4 and 6 show such estimates additionally controlling for all covariates recorded from the contracts as well as seller groups.

	"Third party"	"Theft"	Co-insurance rate
Adverse selection	YES	YES*	NO
Moral hazard	YES*	NO	NO

Table IV.18: Summary of identification of adverse selection and moral hazard. YES* implies results are not robust to all risk measures. For moral hazard in third party higher coverage implies higher loss size, while for adverse selection in theft riskiness, higher riskiness based on three year history of accidents implies higher coverage.

V Adverse selection and interaction with information

A Third party

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.042	-0.044	-0.025	-0.025	-0.321	-0.290
	(0.038)	(0.039)	(0.035)	(0.036)	(0.409)	(0.426)
log(Days insured)	0.089***	0.091***	0.060**	0.062**	0.538***	0.561***
	(0.032)	(0.033)	(0.025)	(0.026)	(0.148)	(0.146)
Male		0.062*		0.049*		0.593*
		(0.032)		(0.028)		(0.347)
Age		-0.002		-0.001		-0.017
		(0.001)		(0.001)		(0.012)
log(Carval)		-0.060		-0.042		-0.405
		(0.038)		(0.033)		(0.250)
Other controls	NO	YES	NO	YES	NO	YES
N	389	388	389	388	389	388
R^2					0.021	0.041
Log Likelihood	-128.001	-122.330	-110.566	-105.994		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.1: Survey respondents, admin claims data (exc biased contracts)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.304***	-0.298***	-0.152**	-0.146**	-0.811	-0.890
	(0.097)	(0.099)	(0.066)	(0.067)	(0.721)	(0.740)
Male		0.072		0.052		1.206**
		(0.089)		(0.053)		(0.599)
Age		-0.009**		-0.004**		-0.043**
		(0.004)		(0.002)		(0.021)
log(Carval)		-0.056		-0.022		-0.646
		(0.096)		(0.054)		(0.573)
Other controls	NO	YES	NO	YES	NO	YES
N	389	388	389	388	389	388
R^2					0.003	0.037
Log Likelihood	-440.594	-434.292	-266.442	-261.227		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.2: Survey respondents, three year history of accidents (exc biased contracts)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.167	0.095	0.171	0.129	0.856	0.677
	(0.233)	(0.182)	(0.154)	(0.142)	(0.735)	(0.730)
Knowledge	0.102**	0.108***	0.095***	0.099***	1.083***	1.182***
	(0.040)	(0.041)	(0.032)	(0.031)	(0.325)	(0.349)
Choosing lower ins*Knowledge	-0.118**	-0.094*	-0.106***	-0.093**	-1.683*	-1.376
	(0.054)	(0.053)	(0.039)	(0.040)	(0.883)	(0.862)
log(Days insured)	0.084***	0.089***	0.055**	0.058**	0.525***	0.558***
	(0.032)	(0.034)	(0.025)	(0.025)	(0.154)	(0.153)
Male		0.052		0.040		0.474
		(0.033)		(0.028)		(0.345)
Age		-0.002		-0.001		-0.016
		(0.001)		(0.001)		(0.012)
log(Carval)		-0.079*		-0.057*		-0.661**
		(0.041)		(0.035)		(0.279)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.038	0.059
Log Likelihood	-121.276	-115.751	-103.372	-98.863		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.3: Interaction with knowledge: Survey respondents, admin claims data

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Informed (Knowledge=1)						
Choosing lower ins	-0.086*	-0.074	-0.065	-0.055	-0.827*	-0.704
	(0.045)	(0.051)	(0.041)	(0.046)	(0.486)	(0.507)
Other controls	NO	YES	NO	YES	NO	YES
N	268	267	268	267	268	267
R^2					0.035	0.070
Log Likelihood	-100.089	-92.038	-86.258	-80.073		
Uninformed (Knowledge=0)						
Choosing lower ins	0.055	$1.414*10^2$	0.064	0.265	0.751	1.199*
	(0.067)	$(7.007*10^2)$	(0.062)	(8.275)	(0.743)	(0.715)
Other controls	NO	YES	NO	YES	NO	YES
N	109	109	109	109	109	109
R^2					0.018	0.093
Log Likelihood	-20.542	-9.569	-16.292	-7.256		

Table V.4: Adverse selection for informed and uninformed survey respondents (admin claims data)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.038	0.003	0.094	0.069	1.461	1.088
	(0.173)	(0.168)	(0.112)	(0.114)	(1.220)	(1.286)
Knowledge	0.165**	0.173**	0.112**	0.123**	1.348**	1.478**
	(0.074)	(0.077)	(0.053)	(0.054)	(0.627)	(0.634)
Choosing lower ins*Knowledge	-0.337***	-0.322***	-0.242***	-0.226***	-3.554**	-3.161**
	(0.091)	(0.097)	(0.069)	(0.077)	(1.411)	(1.486)
log(Days insured)	-0.100***	-0.099***	-0.060**	-0.063**	-0.693**	-0.752**
	(0.035)	(0.036)	(0.027)	(0.028)	(0.328)	(0.328)
Male		0.023		0.044		0.850
		(0.072)		(0.050)		(0.561)
Age		-0.003		-0.002		-0.027
		(0.003)		(0.002)		(0.019)
log(Carval)		-0.062		-0.041		-0.628
		(0.078)		(0.053)		(0.557)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.035	0.058
Log Likelihood	-318.169	-314.262	-225.271	-220.299		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.5: Interaction with knowledge: survey respondents, history prior to the contract

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Informed (Knowledge=1)						
Choosing lower ins	-0.353***	-0.337***	-0.204***	-0.186***	-2.093***	-1.832**
	(0.078)	(0.082)	(0.065)	(0.068)	(0.710)	(0.741)
Other controls	NO	YES	NO	YES	NO	YES
N	268	267	268	267	268	267
R^2					0.034	0.074
Log Likelihood	-233.116	-225.191	-163.130	-155.920		
Uninformed (Knowledge=0)						
Choosing lower ins	0.047	0.128	0.092	0.131	1.497	1.635
	(0.144)	(0.191)	(0.109)	(0.124)	(1.222)	(1.383)
Other controls	NO	YES	NO	YES	NO	YES
N	109	109	109	109	109	109
R^2					0.027	0.079
Log Likelihood	-84.499	-81.504	-62.111	-59.683		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.6: Adverse selection for informed and uninformed survey respondents (history of accidents)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.326	0.561	0.201*	0.265**	-0.156	-0.073
	(0.255)	(0.395)	(0.122)	(0.125)	(0.155)	(0.074)
Knowledge	0.136**	0.143**	0.115***	0.116***	-0.139	-0.117
	(0.054)	(0.056)	(0.041)	(0.041)	(0.138)	(0.114)
Choosing lower ins*Knowledge	-0.218***	-0.258***	-0.176***	-0.188***	0.156	0.063
	(0.060)	(0.078)	(0.039)	(0.036)	(0.155)	(0.067)
log(Days insured)	0.054*	0.060*	0.048*	0.054**	-0.053	-0.044
	(0.030)	(0.031)	(0.025)	(0.026)	(0.053)	(0.044)
Male		-0.066		-0.053		-0.052
		(0.052)		(0.041)		(0.052)
Age		-0.005**		-0.004**		-0.002
		(0.002)		(0.002)		(0.002)
log(Carval)		0.008		-0.0003		-0.003
		(0.053)		(0.043)		(0.013)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.014	0.035
Log Likelihood	-184.367	-177.583	-159.414	-154.031		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.7: Interaction with knowledge: Survey respondents, unclaimed accidents during the contract

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Informed (Knowledge=1)						
Choosing lower ins	-0.151***	-0.151***	-0.125**	-0.119**	0.000	0.000
	(0.054)	(0.056)	(0.050)	(0.052)	(0.000)	(0.000)
Other controls	NO	YES	NO	YES	NO	YES
N	268	267	268	267	268	267
Log Likelihood	-140.854	-133.881	-122.349	-114.853		
Uninformed (Knowledge=0)						
Choosing lower ins	0.164	0.501	0.136	0.225*	-0.181	0.082
	(0.111)	(0.391)	(0.092)	(0.134)	(0.181)	(0.097)
Other controls	NO	YES	NO	YES	NO	YES
N	109	109	109	109	109	109
R^2					0.019	0.077
Log Likelihood	-42.415	-39.142	-36.308	-33.953		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.8: Adverse selection for informed and uninformed survey respondents (unclaimed accidents during the contract)

B Theft

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-1.473e+07	-1.135e+07	-0.653***	-0.656***	-6.623**	-7.025**
	(2.251e+10)	(1.635e+10)	(0.063)	(0.063)	(3.314)	(3.349)
Knowledge	-0.085	0.105	-0.024	0.053	-4.968	-4.615
	(0.280)	(0.197)	(0.206)	(0.180)	(3.490)	(3.534)
Choosing lower ins*Knowledge	4.752e+06	2.310e+06	0.316***	0.303***	7.050**	6.893*
	(7.261e+09)	(3.326e+09)	(0.061)	(0.056)	(3.536)	(3.578)
Male		0.173		0.132		1.437*
		(0.134)		(0.085)		(0.785)
Age		-0.012*		-0.006		-0.002
		(0.007)		(0.004)		(0.031)
Other controls	NO	NO	NO	NO	NO	NO
N	95	95	95	95	95	95
R^2					0.077	0.100
Log Likelihood	-69.161	-66.256	-48.540	-46.263		

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.9: Interaction with knowledge: Survey respondents, three year history of accidents (controlling for gender and age only)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Informed (Knowledge=1)						
Classdummy	-0.261	-0.786*	-0.209	-0.351**	0.427	-0.256
	(0.213)	(0.469)	(0.140)	(0.137)	(1.235)	(1.181)
Other controls	NO	NO	NO	NO	NO	NO
N	77	77	77	77	77	77
R^2					0.001	0.035
Log Likelihood	-64.605	-60.200	-45.767	-43.118		
Uninformed (Knowledge=0)						
Classdummy	-0.750*	-0.926	-0.500**	-0.372	-6.623**	-6.534**
	(0.433)	(1.405)	(0.250)	(0.291)	(3.314)	(3.301)
Other controls	NO	NO	NO	NO	NO	NO
N	18	18	18	18	18	18
R^2					0.437	0.438
Log Likelihood	-4.556	-4.005	-2.773	-2.695		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.10: Adverse selection for informed and uninformed survey respondents (three year history of accidents)

C Coinsurance rate

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.076	0.087	0.015	0.053	0.195	0.691
	(0.297)	(0.301)	(0.121)	(0.117)	(1.516)	(1.468)
Knowledge	0.115	0.092	-0.062	-0.075	-0.578	-0.727
	(0.314)	(0.322)	(0.125)	(0.123)	(1.647)	(1.620)
Choosing lower ins*Knowledge	0.096	0.112	-0.010	0.005	-0.940	-0.789
	(0.411)	(0.424)	(0.162)	(0.160)	(2.122)	(2.094)
log(Days insured)	-0.318***	-0.376***	-0.065	-0.083	-1.011*	-1.379**
	(0.108)	(0.117)	(0.051)	(0.052)	(0.528)	(0.547)
Male		-0.125		-0.004		-0.041
		(0.209)		(0.081)		(1.072)
Age		0.001		-0.001		-0.020
		(0.008)		(0.003)		(0.043)
log(Carval)		0.203		0.052		0.340
		(0.215)		(0.086)		(1.155)
Other controls	NO	YES	NO	YES	NO	YES
N	148	148	148	148	148	148
R^2					0.029	0.102
Log Likelihood	-233.857	-231.829	-94.660	-88.202		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table V.11: Interaction with knowledge: Survey respondents, risk measures for Collision and Third party

VI Adverse selection and heterogeneity in informational value

A Third party

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	3.822	3.975	0.475***	0.456***	5.544*	5.716*
	(4.596)	(4.874)	(0.155)	(0.168)	(3.317)	(3.474)
Knowledge	1.314*	1.393*	0.357***	0.367***	4.574**	4.861***
	(0.761)	(0.825)	(0.070)	(0.065)	(1.783)	(1.817)
Choosing lower ins*Knowledge	-8.135	-8.288	-0.382***	-0.382***	-14.055***	-14.011***
	(15.631)	(16.133)	(0.023)	(0.023)	(4.053)	(4.217)
Choosing lower ins*Age	-0.023**	-0.024**	-0.010	-0.010	-0.091	-0.105
	(0.011)	(0.011)	(0.007)	(0.007)	(0.063)	(0.067)
Knowledge*Age	-0.020***	-0.020***	-0.010**	-0.010**	-0.080*	-0.082*
	(0.006)	(0.007)	(0.004)	(0.004)	(0.042)	(0.042)
Choosing lower ins*Know*Age	0.048**	0.049**	0.026**	0.026**	0.245***	0.253***
	(0.019)	(0.019)	(0.011)	(0.011)	(0.089)	(0.093)
Age	0.012**	0.012**	0.004	0.004	0.023	0.025
	(0.005)	(0.005)	(0.004)	(0.004)	(0.031)	(0.032)
log(Days insured)	-0.107***	-0.106***	-0.064**	-0.067**	-0.728**	-0.787**
	(0.036)	(0.036)	(0.027)	(0.028)	(0.328)	(0.325)
Male		0.005		0.036		0.800
		(0.073)		(0.049)		(0.558)
log(Carval)		-0.063		-0.042		-0.691
		(0.082)		(0.054)		(0.558)
Other controls	NO	YES	NO	YES	NO	YES
N	376	376	376	376	376	376
R^2					0.053	0.073
Log Likelihood	-310.763	-307.769	-220.348	-216.336		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.1: Interaction of informational value with age

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	109.953	104.040	0.584***	0.577***	14.320**	14.761**
	(312.202)	(297.467)	(0.026)	(0.027)	(5.753)	(6.034)
Knowledge	3.983	3.705	0.418***	0.418***	7.616**	7.703**
	(4.274)	(3.981)	(0.034)	(0.034)	(3.318)	(3.259)
Choosing lower ins*Knowledge	-2,262.071	-1,995.413	-0.409***	-0.409***	-27.655***	-27.925***
	(8,795.705)	(7,800.272)	(0.026)	(0.026)	(6.725)	(6.822)
Choosing lower ins*Age	-0.066*	-0.068*	-0.042*	-0.043*	-0.329***	-0.363***
	(0.036)	(0.037)	(0.023)	(0.023)	(0.126)	(0.131)
Knowledge*Age	-0.036***	-0.035***	-0.020**	-0.020**	-0.170**	-0.171**
	(0.012)	(0.012)	(0.008)	(0.008)	(0.083)	(0.081)
Choosing lower ins*Know*Age	0.116**	0.117***	0.071**	0.072***	0.602***	0.632***
	(0.045)	(0.045)	(0.028)	(0.028)	(0.156)	(0.156)
Age	0.024**	0.024**	0.013*	0.013*	0.080	0.088
	(0.011)	(0.011)	(0.007)	(0.007)	(0.070)	(0.068)
log(Days insured)	-0.120***	-0.120***	-0.062**	-0.066**	-0.552	-0.633*
	(0.041)	(0.042)	(0.031)	(0.032)	(0.363)	(0.360)
Male		0.053		0.066		1.245**
		(0.082)		(0.055)		(0.627)
log(Carval)		-0.080		-0.072		-1.049
		(0.092)		(0.063)		(0.646)
Other controls	NO	YES	NO	YES	NO	YES
N	296	296	296	296	296	296
R^2					0.060	0.099
Log Likelihood	-248.013	-244.529	-175.113	-170.793		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.2: Interaction of information value with age excluding extreme age values (top and bottom 10%)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	0.355	0.305	0.245	0.206	2.654	2.431
	(0.401)	(0.382)	(0.166)	(0.172)	(1.940)	(2.021)
Knowledge	0.169	0.184	0.084	0.097	0.323	0.563
	(0.113)	(0.113)	(0.083)	(0.083)	(0.949)	(0.964)
Choosing lower ins*Knowledge	-0.624**	-0.591**	-0.347***	-0.338***	-6.706***	-6.312***
	(0.279)	(0.252)	(0.042)	(0.050)	(2.156)	(2.249)
Choosing lower ins*Exper	-0.016	-0.017	-0.008	-0.007	-0.039	-0.052
	(0.014)	(0.014)	(0.009)	(0.009)	(0.056)	(0.059)
Knowledge*Exper	-0.0003	-0.001	0.003	0.003	0.085*	0.079
	(0.008)	(0.008)	(0.006)	(0.006)	(0.049)	(0.049)
Choosing lower ins*Know*Exper	0.044*	0.043*	0.024	0.022	0.192*	0.194*
	(0.025)	(0.025)	(0.015)	(0.015)	(0.108)	(0.113)
Exper	-0.0001	0.003	-0.003	-0.002	-0.067**	-0.055
	(0.007)	(0.008)	(0.005)	(0.005)	(0.028)	(0.040)
log(Days insured)	-0.100***	-0.100***	-0.060**	-0.063**	-0.683**	-0.737**
	(0.035)	(0.036)	(0.027)	(0.028)	(0.331)	(0.330)
Male		0.008		0.035		0.697
		(0.076)		(0.052)		(0.592)
log(Carval)		-0.077		-0.052		-0.830
		(0.081)		(0.055)		(0.562)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.054	0.075
Log Likelihood	-316.150	-312.295	-222.695	-218.220		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.3: Interaction with driving experience

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.209	-0.234	-0.158	-0.165	-0.972	-1.053
	(0.198)	(0.190)	(0.173)	(0.172)	(2.075)	(2.223)
Knowledge	0.109	0.147	0.077	0.104	1.707**	2.165***
	(0.128)	(0.128)	(0.089)	(0.088)	(0.776)	(0.792)
Choosing lower ins*Knowledge	-0.224	-0.203	-0.116	-0.091	-1.812	-1.629
	(0.207)	(0.228)	(0.221)	(0.238)	(2.223)	(2.379)
Choosing lower ins*Distance	0.008	0.008	0.006	0.006	0.056**	0.050*
	(0.005)	(0.005)	(0.004)	(0.005)	(0.028)	(0.028)
Knowledge*Distance	0.004	0.004	0.002	0.002	0.003	-0.001
	(0.004)	(0.004)	(0.002)	(0.002)	(0.008)	(0.008)
Choosing lower ins*Know*Distance	-0.008	-0.008	-0.006	-0.006	-0.052*	-0.048
	(0.007)	(0.007)	(0.005)	(0.005)	(0.030)	(0.030)
Distance	-0.005	-0.005	-0.003	-0.003	-0.006*	-0.006
	(0.003)	(0.004)	(0.002)	(0.002)	(0.004)	(0.005)
log(Days insured)	-0.100***	-0.100***	-0.052*	-0.056**	-0.534	-0.643*
	(0.035)	(0.035)	(0.027)	(0.028)	(0.333)	(0.334)
Male		0.038		0.049		0.911
		(0.073)		(0.051)		(0.582)
log(Carval)		-0.119		-0.047		-0.681
		(0.089)		(0.059)		(0.577)
Other controls	NO	YES	NO	YES	NO	YES
N	343	342	343	342	343	342
R^2					0.055	0.093
Log Likelihood	-264.698	-257.228	-194.519	-187.350		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.4: Interaction of information value with daily average distance

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.052	-0.195	0.065	-0.027	2.090	1.087
	(0.528)	(0.406)	(0.385)	(0.367)	(3.711)	(3.524)
Knowledge	0.201	0.191	0.208*	0.191	2.704*	2.507
	(0.197)	(0.203)	(0.124)	(0.129)	(1.474)	(1.562)
Choosing lower ins*Knowledge	-0.312	-0.189	-0.278	-0.206	-4.467	-3.140
	(0.309)	(0.496)	(0.189)	(0.297)	(4.111)	(3.884)
Choosing lower ins*Distance	0.002	0.005	-0.00001	0.002	-0.014	-0.001
	(0.013)	(0.013)	(0.008)	(0.008)	(0.075)	(0.068)
Knowledge*Distance	-0.0002	0.001	-0.003	-0.002	-0.023	-0.013
	(0.006)	(0.006)	(0.004)	(0.004)	(0.036)	(0.038)
Choosing lower ins*Know*Distance	-0.002	-0.007	0.001	-0.002	0.002	-0.020
	(0.021)	(0.021)	(0.012)	(0.012)	(0.083)	(0.078)
Distance	-0.003	-0.005	-0.0004	-0.002	0.005	-0.007
	(0.006)	(0.006)	(0.003)	(0.003)	(0.026)	(0.028)
log(Days insured)	-0.074*	-0.077*	-0.030	-0.035	-0.301	-0.418
	(0.040)	(0.041)	(0.031)	(0.031)	(0.364)	(0.371)
Male		0.083		0.068		0.984
		(0.083)		(0.059)		(0.685)
log(Carval)		-0.068		-0.017		-0.208
		(0.103)		(0.069)		(0.685)
Other controls	NO	YES	NO	YES	NO	YES
N	255	255	255	255	255	255
R^2					0.045	0.078
Log Likelihood	-202.332	-195.800	-151.577	-145.686		

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.5: Interaction of info value with distance (excluding extreme top and bottom 10%)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.097	-0.123	0.023	0.018	1.432	1.093
	(0.193)	(0.187)	(0.140)	(0.141)	(1.363)	(1.499)
Knowledge	0.145	0.149	0.095	0.106	2.123***	2.239***
	(0.108)	(0.109)	(0.074)	(0.073)	(0.764)	(0.765)
Choosing lower ins*Knowledge	-0.087	-0.060	-0.055	-0.035	-2.172	-1.809
	(0.240)	(0.261)	(0.155)	(0.162)	(1.852)	(1.933)
Choosing lower ins*High income	0.474	0.477	0.364	0.303	2.679	2.528
	(0.752)	(0.802)	(0.275)	(0.304)	(2.850)	(3.013)
Knowledge*High income	-0.040	-0.016	-0.061	-0.046	-2.312	-2.070
	(0.177)	(0.180)	(0.114)	(0.115)	(1.443)	(1.432)
Choosing lower ins*Know*High income	-0.450***	-0.452***	-0.338***	-0.337***	-5.460*	-5.418
	(0.050)	(0.052)	(0.031)	(0.032)	(3.171)	(3.370)
High income	0.091	0.087	0.089	0.079	2.504**	2.333*
	(0.167)	(0.169)	(0.108)	(0.108)	(1.226)	(1.230)
log(Days insured)	-0.102***	-0.097***	-0.059**	-0.063**	-0.623*	-0.650*
	(0.037)	(0.038)	(0.028)	(0.029)	(0.356)	(0.358)
Male		0.014		0.053		0.680
		(0.075)		(0.051)		(0.601)
log(Carval)		-0.074		-0.043		-0.826
		(0.087)		(0.059)		(0.613)
Other controls	NO	YES	NO	YES	NO	YES
N	349	348	349	348	349	348
R^2					0.062	0.080
Log Likelihood	-288.096	-284.510	-201.029	-196.450		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.6: Interaction of informational value with income: high vs low income

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.418	-0.429	-0.205	-0.219	-1.373	-1.648
	(0.269)	(0.274)	(0.143)	(0.138)	(1.280)	(1.246)
Knowledge	0.124	0.148	0.115	0.129	1.720*	1.914*
	(0.138)	(0.138)	(0.093)	(0.093)	(1.036)	(1.108)
Choosing lower ins*Knowledge	0.301	0.298	0.092	0.106	-0.044	-0.008
	(0.862)	(0.861)	(0.294)	(0.294)	(1.833)	(1.792)
Choosing lower ins*Higher educ	2.702	2.540	0.498***	0.491***	5.050**	4.915**
	(3.484)	(3.332)	(0.157)	(0.158)	(2.119)	(2.225)
Knowledge*Higher educ	0.068	0.042	-0.007	-0.015	-0.544	-0.660
	(0.191)	(0.195)	(0.127)	(0.128)	(1.307)	(1.420)
Choosing lower ins*Know*Higher educ	-0.484***	-0.467***	-0.322***	-0.318***	-6.149**	-5.583**
	(0.185)	(0.175)	(0.056)	(0.062)	(2.625)	(2.756)
Higher educ	-0.061	-0.038	-0.004	0.007	0.234	0.442
	(0.180)	(0.182)	(0.113)	(0.115)	(1.069)	(1.181)
log(Days insured)	-0.104***	-0.103***	-0.065**	-0.068**	-0.742**	-0.798**
	(0.035)	(0.036)	(0.027)	(0.028)	(0.338)	(0.340)
Male		0.018		0.040		0.818
		(0.074)		(0.050)		(0.558)
Age		-0.003		-0.002		-0.027
		(0.003)		(0.002)		(0.019)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.052	0.075
Log Likelihood	-315.340	-311.564	-221.960	-217.065		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.7: Interaction of info value with higher education dummy

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-1.823e+05	-1.637e+04	-0.401***	-0.402***	-27.717	-19.161
	(2.261e+06)	(2.059e+05)	(0.023)	(0.023)	(27.085)	(30.029)
Knowledge	-0.163	-0.035	-0.556***	-0.547***	12.757	13.249
	(3.408)	(2.884)	(0.024)	(0.025)	(17.911)	(19.009)
Choosing lower ins*Knowledge	1.695e+16	8.729e+14	0.618***	0.617***	46.273	40.663
	(3.574e+17)	(1.852e+16)	(0.023)	(0.023)	(35.521)	(37.391)
Choosing lower ins*log(Carval)	0.408	0.343	0.334	0.244	1.843	1.281
	(0.328)	(0.334)	(0.218)	(0.215)	(1.719)	(1.905)
Knowledge*log(Carval)	0.022	0.014	0.114	0.107	-0.703	-0.736
	(0.172)	(0.174)	(0.118)	(0.117)	(1.117)	(1.187)
Choosing lower ins*Know*log(Carval)	-1.040*	-0.963*	-0.716**	-0.628*	-3.109	-2.730
	(0.562)	(0.566)	(0.344)	(0.334)	(2.214)	(2.336)
log(Carval)	-0.070	-0.072	-0.136	-0.120	-0.059	-0.030
	(0.159)	(0.170)	(0.109)	(0.114)	(0.971)	(1.143)
log(Days insured)	-0.099***	-0.100***	-0.056**	-0.061**	-0.665**	-0.753**
	(0.036)	(0.036)	(0.027)	(0.027)	(0.322)	(0.327)
Male		0.021		0.041		0.887
		(0.072)		(0.050)		(0.562)
Age		-0.002		-0.002		-0.023
		(0.003)		(0.002)		(0.020)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.046	0.064
Log Likelihood	-315.757	-312.504	-222.137	-218.092		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.8: Interaction of informational value with log(Carval)

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-8.902e+10	-2.541e+10	-0.407***	-0.409***	-114.048	-99.415
	(2.214e+12)	(6.502e+11)	(0.027)	(0.027)	(82.221)	(85.339)
Knowledge	-3.674	-0.071	-0.554***	-0.551	-25.719	-13.528
	(54.072)	(6.878)	(0.027)	(0.991)	(42.764)	(42.229)
Choosing lower ins*Knowledge	1.390e+19	5.219e+17	0.596***	0.595***	146.003	123.692
	(5.125e+20)	(1.947e+19)	(0.027)	(0.027)	(95.098)	(98.434)
Choosing lower ins*log(Carval)	0.806	0.768	0.631	0.549	7.234	6.272
	(0.703)	(0.725)	(0.470)	(0.466)	(5.101)	(5.297)
Knowledge*log(Carval)	0.081	0.016	0.307	0.225	1.678	0.913
	(0.388)	(0.392)	(0.260)	(0.256)	(2.656)	(2.623)
Choosing lower ins*Know*log(Carval)	-1.308	-1.216	-1.033*	-0.924	-9.359	-7.919
	(1.045)	(1.059)	(0.619)	(0.612)	(5.893)	(6.106)
log(Carval)	-0.045	-0.024	-0.157	-0.090	-1.046	-0.658
	(0.349)	(0.365)	(0.229)	(0.235)	(2.186)	(2.220)
log(Days insured)	-0.101**	-0.103**	-0.053*	-0.060*	-0.689*	-0.784**
	(0.043)	(0.044)	(0.031)	(0.032)	(0.392)	(0.397)
Male		0.039		0.050		0.770
		(0.087)		(0.058)		(0.667)
Age		-0.0001		-0.0004		-0.005
		(0.003)		(0.002)		(0.025)
Other controls	NO	YES	NO	YES	NO	YES
N	280	279	280	279	280	279
R^2					0.047	0.077
Log Likelihood	-245.188	-241.718	-167.650	-163.979		

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.9: Interaction of info value with log(Carval), excluding extreme values

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	2.638	2.306	0.235	0.195	0.463	-0.112
	(7.447)	(6.714)	(0.673)	(0.689)	(7.329)	(7.282)
Knowledge	-0.134	-0.172	-0.448***	-0.451***	-8.903**	-9.594**
	(0.679)	(0.731)	(0.169)	(0.155)	(3.966)	(3.940)
Choosing lower ins*Knowledge	-0.100	-0.008	0.415	0.468	3.200	3.679
	(1.221)	(1.551)	(0.680)	(0.524)	(8.651)	(8.609)
Choosing lower ins*Riskloving	-0.245	-0.241	-0.043	-0.039	0.208	0.254
	(0.309)	(0.313)	(0.188)	(0.188)	(2.032)	(2.013)
Knowledge*Riskloving	0.082	0.093	0.182*	0.190*	2.819***	3.041***
	(0.154)	(0.157)	(0.106)	(0.106)	(1.042)	(1.040)
Choosing lower ins*Know*Riskloving	-0.103	-0.121	-0.195	-0.209	-1.776	-1.797
	(0.444)	(0.450)	(0.255)	(0.254)	(2.372)	(2.352)
Riskloving	0.015	0.010	-0.085	-0.085	-1.634**	-1.724**
	(0.137)	(0.140)	(0.091)	(0.090)	(0.816)	(0.809)
log(Days insured)	-0.105***	-0.106***	-0.063**	-0.069**	-0.706**	-0.794**
	(0.035)	(0.036)	(0.027)	(0.028)	(0.328)	(0.327)
Male		0.027		0.052		0.958*
		(0.072)		(0.049)		(0.559)
Age		-0.003		-0.002		-0.026
		(0.003)		(0.002)		(0.020)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.054	0.079
Log Likelihood	-316.535	-312.527	-222.345	-217.062		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.10: Interaction of info value with risk aversion

	Acc frequency		Prob accidents		Claim size	
	Poisson		*logistic*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
Choosing lower ins	-0.006	-0.026	0.068	0.059	1.012	0.704
	(0.181)	(0.179)	(0.125)	(0.126)	(1.367)	(1.436)
Knowledge	0.131	0.137	0.104*	0.113*	1.313*	1.396*
	(0.088)	(0.091)	(0.063)	(0.063)	(0.751)	(0.764)
Choosing lower ins*Knowledge	-0.269**	-0.260**	-0.189*	-0.180*	-2.834*	-2.570
	(0.126)	(0.131)	(0.100)	(0.104)	(1.645)	(1.704)
Choosing lower ins* Friend	0.226	0.146	0.140	0.070	2.265	1.778
	(0.586)	(0.522)	(0.292)	(0.284)	(3.011)	(2.985)
Knowledge* Friend	0.166	0.180	0.035	0.045	0.101	0.228
	(0.256)	(0.263)	(0.137)	(0.137)	(1.352)	(1.339)
Choosing lower ins*Know* Friend	-0.336**	-0.319*	-0.251**	-0.227	-3.017	-2.382
	(0.148)	(0.170)	(0.114)	(0.148)	(3.316)	(3.308)
Friend	-0.092	-0.095	0.004	0.005	-0.504	-0.643
	(0.168)	(0.170)	(0.115)	(0.115)	(1.080)	(1.070)
log(Days insured)	-0.100***	-0.099***	-0.060**	-0.063**	-0.702**	-0.768**
	(0.035)	(0.036)	(0.027)	(0.027)	(0.326)	(0.326)
Male		0.021		0.041		0.888
		(0.073)		(0.050)		(0.565)
Age		-0.003		-0.002		-0.026
		(0.003)		(0.002)		(0.020)
Other controls	NO	YES	NO	YES	NO	YES
N	377	376	377	376	377	376
R^2					0.038	0.060
Log Likelihood	-317.361	-313.462	-224.389	-219.499		

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table VI.11: Interaction of info value with friendship with the seller