Asymmetric Information in Automobile Insurance: New Evidence from Telematics Data*

Alexander Muermann† Daniela Straka†

December 2010

We provide novel insights into the effects of private information in automobile insurance. Our analysis is based on telematic data of insured cars which includes detailed information about driving behavior that is unobservable to the insurance company. We find that private information about driving behavior has significant and counteracting effects on the choice of third-party liability and first-party insurance coverage. Yet, there is no residual correlation between the level of insurance coverage and risk. These results suggest overlapping and offsetting effects of private information based on risk preferences and driving behavior.

*We are grateful to Alois Geyer and Christian Laux for their helpful comments and suggestions.
†Department of Finance, Accounting and Statistics, Vienna University of Economics and Business, Heiligenstaedter Str. 46-48, A - 1190 Vienna, AUSTRIA, alexander.muermann@wu.ac.at – daniela.straka@wu.ac.at

1 Introduction

This paper provides new insights into the relevance of private information for contract choice in insurance markets based on a telematics data set of insured cars which is inaccessible to the insurance company. The data set contains detailed information about driving behavior and is recorded by a telematics device which is installed in the insured cars. While the insurance company uses the aggregate distance driven for premium calculation, it contractually refrains itself from accessing any other telematic data.[1] We also have access to the corresponding insurance data set which we are able to link to the telematics data set on the car level. The combination of insurance and telematic data and the fact that most of the telematic data is unobservable to the insurance company allows us to test whether private information about driving behavior is relevant and, if so, how it is linked to the policyholder's choice of insurance contract and the conditional loss distribution.

Controlling for all characteristics that are observable to the insurance company, we find two types of driving behavior to be significantly linked to contract choice and a subsequent downgrade in the Bonus-Malus class: the number of car rides a policyholder undertakes and average speeding above legal speed limits. The number of car rides and average speeding above legal speed limits are both negatively related to the level of third-party liability coverage. In contrast, the number of car rides is positively related to the level of first-party insurance coverage. Last, the number of car rides is positively related to a downgrade in the Bonus-Malus class.

Our results suggest overlapping and counteracting effects of asymmetric information based on risk preferences and driving behavior. The negative relation of the number of car rides and of average speeding to the level of liability coverage indicate a selection and incentive effect based on hidden risk preferences. More risk-averse policyholders both purchase more liability coverage and invest more in risk reduction by undertaking fewer car rides and speeding on average less above legal speed limits. The positive relation between the number of car rides to the level of first-party coverage and to a downgrade in the Bonus-Malus class suggest a selection and/or incentive effect based on driving behavior. Policyholders who undertake more car rides purchase more first-party insurance coverage and are more likely to be downgraded in their Bonus-Malus class.

[1] The production and installation of the hardware into the cars as well as the collection and management of the telematic data is carried out by an independent telematics company.

These counteracting effects of asymmetric information pose a challenge for empirical tests based on the residual correlation between the level of insurance coverage and ex-post realizations of risk. Failing to reject the null hypothesis of zero residual correlation could either indicate the absence of private information or that multiple, counteracting effects of private information cancel each other out with respect to the residual correlation. We apply the standard test based on residual correlation to our insurance data and, in fact, cannot reject the null hypothesis of zero residual correlation. Our result confirms that the absence of residual correlation between the level of insurance coverage and ex-post realizations of risk is not sufficient to conclude that private information is absent or irrelevant. We provide empirical evidence that driving behavior is related to changes in the risk class and that private information about driving behavior influences contract choice. But, the overlapping and counteracting effects result in no significant residual correlation between insurance coverage and ex-post realizations of risk.

Most of the empirical literature on asymmetric information in insurance markets analyzes insurance data and estimates the sign of the correlation between the level of insurance coverage and ex post realizations of risk. The classical models both of adverse selection and moral hazard (Arrow, 1963; Pauly, 1974; Rothschild and Stiglitz, 1976; Harris and Raviv, 1978; Holmstrom, 1979; Shavell, 1979) predict a positive correlation which is confirmed in the health insurance market (Cutler and Reber, 1998; Cutler and Zeckhauser, 1998) and in the annuity market (Finkelstein and Poterba, 2002, 2004). However, there is also evidence for a negative correlation between the level of insurance coverage and ex post realizations of risk in the markets for life insurance (Cawley and Philipson, 1999) and for Medigap insurance (Fang et al., 2008). Moreover, no statistically significant correlation has been found in automobile insurance (Chiappori and Salanié, 2000; Dionne et al., 2001; Cohen, 2005) and in long-term care insurance (Finkelstein and McGarry, 2006).[2] We refer to Cohen and Siegelmann (2010) for a review of the empirical literature on asymmetric information in insurance markets.

Our result of offsetting effects of asymmetric information based on risk preferences and driving behavior shows that it is difficult to draw unambiguous conclusions about the effects of private information based on the residual correlation between the level of insurance coverage and risk. This is consistent with the literature that examines the effect of hidden risk preferences. Chiap-

[2] Puelz and Snow (1994) did find a positive relation between coverage and risk. Their result, however, was subsequently challenged by Chiappori and Salanié, 2000, and Dionne et al., 2001. While Cohen (2005) did not find any correlation for beginning drivers, she did find a statistically significant positive relation for experienced drivers.

pori et al. (2006) examine theoretically the extent to which adverse selection and moral hazard models can be generalized while still predicting a positive correlation between level of insurance coverage and risk. They emphasize that hidden degree of risk aversion can be pivotal for violating the prediction of positive correlation. In fact, de Meza and Webb (2001) show in a theoretical model that a separating equilibrium with a negative relation between coverage and risk exists if hidden information about the degree of risk aversion is combined with hidden investment in risk reduction. Finkelstein and Poterba (2006) also argue that if asymmetric information is present on multiple characteristics, including the degree of risk aversion, then the result of rejecting (not rejecting) the hypothesis of non-dependence between the level of insurance coverage and risk may not be indicative of the existence (absence) of asymmetric information. Cohen and Einav (2007) develop a structural model which accounts for unobserved heterogeneity in both risk and risk aversion. By using a large data set of an Israeli insurance company, they find that unobserved heterogeneity in risk aversion is much larger than unobserved heterogeneity in risk.

We address the problem of potentially overlapping effects by providing direct evidence of the existence and effect of private information based on the telematics data set which is unobservable to the insurance company. This relates our work to Finkelstein and Poterba (2006) who propose an empirical test based on "unused observables," i.e. on characteristics which are observable to the insurance company but are not used for pricing, either voluntarily or for legal reasons. They argue that if those characteristics are significantly related to contract choice and risk, then this is direct evidence of relevant private information which is not confounded by hidden information on risk preferences. In their study of the UK annuity market, they use postcode information which is collected by the insurance company but not used for pricing. They find that the inhabitants' socio-economic characteristics of different postcode areas are correlated with both survival probability and choice of insurance coverage. Similarly, Saito (2006) uses postcode information which is collected but not used by insurance companies for pricing in automobile insurance. The author rejects the hypothesis that policyholders who live in high accident probability regions are more likely to purchase insurance. A potential problem with unused but observable data is that the information, although not used in pricing, might be used in other types of underwriting activities by the insurance company. For example, policyholders who observably differ in their underlying risk might be offered different contracts, might be scrutinized differently in the claims settlement process, or might face different cancellation policies. In that case, a significant relation between the "unused observable" and contract choice might reflect rather those different underwriting policies than an effect of private information. In our paper, we have access to data which provides us with information which is *unobservable* to the insurance company. Thus, the insurance company is not able to condition any type of underwriting activity on that information.

The utilization of information which is unobservable to the insurance company relates our paper to work that is based on survey data. In addition to information on insurance coverage and realization of risks, some surveys include information about the interviewees which is unobservable to the insurance company. Finkelstein and McGarry (2006) use individual-level survey data and show that individuals' self-reported beliefs of entering a nursing home is positively related to both subsequent nursing home use and insurance coverage. Despite the existence of this risk-based selection, actual nursing home use and insurance coverage is not positively correlated. The authors explain this fact by providing evidence that the risk-based selection is offset by a selection based on heterogeneous degrees of risk aversion. Fang et al. (2008) also use individual-level survey data to examine the reasons for the significant and negative correlation between insurance coverage and medical expenditure in the Medigap insurance market. They show that cognitive ability rather than risk preferences is the essential factor explaining this negative relation. A potential problem with using survey data is that responses to survey questions can be biased, in particular if they relate to self-reported probabilities of future events. Examples include the anchoring bias of unfolding bracket questions (Hurd et al., 1998; Hurd, 1999) and problems of focal responses (Gan et al., 2005). These survey response biases could then partially explain the relation between self-reported information and both contract choice and realized risk. In contrast, we can rely on data which contains private information about real decisions and behavior.

The paper is structured as follows. In Section 2, we present the telematics data and insurance data sets and their descriptive statistics. We introduce the indices for driving behavior and the econometric model in Section 3 and present and discuss the results in Section 4. In Section 5, we examine extensions and check the robustness of our results.

2 Data

The insurance company introduced a pay-as-you-drive insurance contract. Cars insured under this contract are equipped with a telematics device which uses GPS. The pricing of this pay-as-you-drive contract is based on the aggregate distance driven - fewer kilometers driven imply a lower premium - and on the road type used. The company distinguishes between three road types: urban, country road, and motorway. Kilometers driven on country roads and motorways are scaled down by a coefficient of 0.8. Furthermore, policyholders get a discount on the premium of full comprehensive insurance coverage. In addition to the pay-as-you-drive feature, the telematics device is equipped with an emergency device and a crash sensor. If activated, either by the car driver or in case of an accident, an emergency signal is sent to the help desk of the insurance company. The help desk will then try to contact the policyholder and call the police and ambulance if needed or if the policyholder cannot be reached. An additional benefit of the telematics device is that stolen cars can be tracked via GPS. Policyholders have to pay a one-time fee for the installation of the telematics device and a monthly fee for the safety services.

2.1 Telematic Data

An independent telematics company develops the hardware and collects and manages the telematic data. Each data point includes date, time, GPS-coordinates, direction of driving, actual speed, distance to the last data point, ignition status of the engine, and road type (urban, country road, or motorway). A data point is recorded when the engine is started, after approximately every two kilometers driven, and when the engine is switched off. We have access to this data set for 2,340 cars for a period of 3 months, from February 1st, 2009 to April 30th, 2009, which includes 3.7 million individual data points.

For our analysis we restrict our data set to completed car rides, i.e. for which the engine was switched on and in some distance switched off. We exclude car rides with unrealistically high values for speed (above 200 km/h) and distances between data points (above the 99.9% quantile) which indicate a connection failure with the GPS satellite. These exclusions leave us with 3.15 million data points. Table 1 displays the descriptive statistics of the telematic data.

Table 1: DESCRIPTIVE STATISTICS TELEMATIC DATA

	urban	country road	motorway	total
Number of cars				2,340
Number of car rides				537,181
Number of data points	1,717,049	686,042	744,542	3,147,633
Average speed in km/h (mph)	47.72	73.87	113.22	78.03
	(29.65)	(45.90)	(70.35)	(48.49)
Distance driven in km (mi)	2,041,466	1,195,018	1,567,140	4,803,624
	(1,268,508)	(742,550)	(973,776)	(2,984,834)
Avg. distance per ride in km (mi)				8.94
				(5.56)

The insurance company has access only to the telematic data that is necessary for the pricing of the pay-as-you-drive contract, i.e., to the aggregate distance per road type. The insurer contractually refrains itself from accessing any other telematic data. The telematic data set thus provides us with detailed private information on driving behavior which is inaccessible for the insurance company.

2.2 Insurance Data

For all privately insured cars in the telematics data set we have the corresponding data of the insurance contract which we can link to the telematic data on the car level. We thus exclude all corporate cars. The insurance data set comprises all the information used for pricing of the pay-as-you-drive insurance policies in February 2009. Additionally, we have an update of the contract data set for February 2010. We restrict the data set to those cars which are still insured under the pay-as-you-drive contract after one year. Moreover, we only include cars with more than 4 kW (5.4 HP). This leaves us with 1849 cars and insurance contracts for our analysis.

For each contract, the insurance data contains the following information:

1. *Car-related information:* year of construction, brand, engine power, and value of the car
2. *Policyholder-related information:* age, gender, and postal code (urban / rural)
3. *Bonus-Malus class:* Premiums for third-party liability insurance are based on a Bonus-Malus scheme. The national insurance association monitors a Bonus-Malus record for each nationwide registered car owner, which is accessible to all insurance companies.
4. *Downgrade in Bonus-Malus class:* We use downgrades in the Bonus-Malus record between February 2009 and February 2010 to proxy for risk.

5. *Coverage of first-party insurance:* The insurance company offers three levels of first-party coverage: none, comprehensive insurance (covers losses from vandalism, theft, weather etc.), and full comprehensive insurance (in addition including at-fault collision losses).[3]

6. *Coverage of third-party liability insurance:* The insurance company offers two levels of third-party liability coverage which are both in excess of the level of coverage mandated by the insurance law: € 10 million or € 15 million.

Table 2 gives an overview of the insurance data.

Table 2: DESCRIPTIVE STATISTICS INSURANCE DATA

	Mean				
	total	none/compr.	full compr.	liab. 10m	liab. 15m
car's characteristics:					
years since construction	3.47	6.74	1.92	3.68	2.99
kW (HP)	87.09	83.52	88.78	86.57	88.3
	(116.74)	(111.96)	(119.01)	(116.05)	(118.36)
value of car in €	26,709	26,204	27,023	26,656	26,835
policyholder's characteristics:					
age in years	48.67	48.16	48.91	48.13	49.91
male	0.61	0.61	0.61	0.39	0.38
urban	0.44	0.42	0.45	0.45	0.42
BM (Bonus-Malus class)	0.52	0.55	0.51	0.52	0.51
downgrade BM class in %	7.6	9.7	6.6	7.9	7.0

Notes: Column 2 "none/compr." includes contracts with no first-party insurance coverage or comprehensive coverage; Column 3 "full compr." includes contracts with comprehensive coverage and at-fault collision; Bonus-Malus class gives the scaling factor for the premium of liability coverage.

3 Empirical Approach

3.1 Driving Behavior

We investigate two types of individual driving behavior utilizing the information contained in our telematics data set: average speeding above legal speed limits and the number of car rides. The speeding index is given by

$$AvgSpeeding = \frac{\sum_j \sum_{i \in \Delta_n} (v_{ij} - u_j)}{n} \tag{1}$$

[3]We do not use the information on deductible choice since the standard deductible of € 300 is chosen by more than 99% of all policyholders.

where j is road type (urban, country, motorway), u_j is the countrywide legal speed limit for road type j in km/h, $i = 1, ..., n$ is a data point, v_{ij} is the speed of the car at data point i on road type j, and $\Delta_n = \{i = 1, \ldots, n | v_{ij} > u_j\}$ is the set of data points where the speed of the car is above the legal speed limit.[4]

The second index *#Rides* is the number of car rides driven between February 1st, 2009 and April 30th, 2009. We define a car drive if the engine is switched on, a distance is driven, and the engine is switched off.

We derive both indices for each car in our data set. The descriptive statistics for these two indices are given in Table 3.

Table 3: DESCRIPTIVE STATISTICS INDICES

		N			Mean AvgSpeeding			Mean #Rides		
		0	1	total	0	1	total	0	1	total
Coverage	first-party	562	1254	1849	3.05	3.19	3.15	217	255	242
	third-party	1293	556	1849	3.2	3.02	3.15	246	236	242
Risk		1708	141	1849	3.15	3.16	3.15	238	293	242

Notes: first-party coverage is 1 for full comprehensive insurance and 0 otherwise; liability insurance is set to 0, if € 10m are covered and is 1, if coverage is€ 15m.

3.2 Econometric Model

We test for the effect of private information on contract choice and its relation to risk by applying the econometric model suggested by Finkelstein and Poterba (2006). This model extends the econometric model of Chiappori and Salanié (2000) by incorporating information which is observable but not used by the insurance company. More specifically, Chiappori and Salanié (2000) propose the following bivariate probit model for insurance coverage and risk

$$Coverage = 1(X\beta + \varepsilon > 0) \quad (2)$$

$$Risk = 1(X\gamma + \eta > 0) \quad (3)$$

where X is the vector of all risk classifying variables used by the insurance company. Under the null hypothesis of no asymmetric information the correlation ρ of the error terms ϵ and η is zero.

[4] We note that we do not have data on the legal speed limit at each data point. By using the countrywide legal speed limit for each road type we thus underestimate the extent to which drivers speed.

Rejecting the null hypothesis would thus indicate the existence of private information. A statistically significant, positive correlation coefficient is consistent with the classical models of adverse selection and moral hazard (Arrow, 1963; Pauly, 1974; Rothschild and Stiglitz, 1976; Harris and Raviv, 1978; Holmstrom, 1979; Shavell, 1979) with asymmetric information about one parameter of the loss distribution. Chiappori et al. (2006) show that this prediction can be extended to general settings, including, for example, heterogeneous preferences and multidimensional hidden information linked with hidden action, if the insurers' profits are nonincreasing in the level of coverage. However, if the insurers' profits do not satisfy this condition, then Chiappori et al. (2006) point out that the prediction about the positive relation between the level of insurance coverage and risk might no longer hold if the degree of risk aversion is private information.

Finkelstein and Poterba (2006) propose the following extension of Chiappori and Salanié (2000)

$$Coverage = 1(X\beta_1 + Y\beta_2 + \varepsilon > 0) \tag{4}$$

$$Risk = 1(X\gamma_1 + Y\gamma_2 + \eta > 0) \tag{5}$$

where Y is the private information which is observable but not used by the insurance company. Under the null hypothesis of no asymmetric information we have $\beta_2 = 0$ and $\gamma_2 = 0$. The benefit of this model extension is that the rejection of the null hypothesis directly proves that private information contained in Y is relevant for contract choice and/or risk, independent of the type of asymmetric information and independent of how the insurers' profits relate to the level of coverage. This model captures just as well our situation in which the information Y is not observed by the insurance company but accessible to the econometrician.

We apply this econometric model and define the variables as follows. Since policyholders can choose the level of coverage along the two dimensions, first-party and liability coverage, we run two separate bivariate probit models. For first-party coverage, we set $Coverage = 1$ if the contract covers at-fault losses (full comprehensive insurance) and $Coverage = 0$ otherwise. For liability coverage, we set $Coverage = 1$ if the upper limit is € 15m and $Coverage = 0$ if the upper limit is € 10m. The dependent variable $Risk$ is set to 1 if the Bonus-Malus level worsened within the following year and is set to 0 otherwise.

X is the set of variables which the insurance company observes and uses for the pricing of the contract (see Section 2.2). In addition, we also include the aggregate distance driven by the policyholder since this is the part of the telematic data which the insurance company observes and uses for setting the premium. Last, we also add the coverage information that is not used as a dependent variable, e.g., in the bivariate model where the dependent variable *Coverage* specifies first-party coverage, the level of liability coverage is included in the set of observed variables X.

For the set Y of variables that are not observed by the insurance company we include the two indices *AvgSpeeding* and *#Rides* that characterize driving behavior and are constructed from the telematics data set (see Section 3.1). We thus apply the following bivariate probit model

$$Coverage = 1(X\beta_1 + \beta_2 AvgSpeeding + \beta_3 \#Rides + \varepsilon > 0) \tag{6}$$

$$Risk = 1(X\gamma_1 + \gamma_2 AvgSpeeding + \gamma_3 \#Rides + \eta > 0) \tag{7}$$

and test the null hypothesis of no relevant private information, i.e. we test the hypothesis $\beta_2 = \beta_3 = \gamma_2 = \gamma_3 = 0$.[5]

To compare the potential conclusions drawn from the two models, we also test for the correlation coefficient ρ of the error terms ϵ and η both excluding the two indices *AvgSpeeding* and *#Rides*, i.e. equations (2) and (3), and including the two indices, equations (6) and (7).

4 Results and Discussion

4.1 First-Party Insurance Coverage, Driving Behavior, and Risk

In Table 4, we present the results of the bivariate probit model, equations (6) and (7), where the dependent variable *Coverage* in equation (6) is first-party insurance coverage. The first column of Table 4 reports the correlation coefficient between the error terms ϵ and η of the bivariate model in which we do not include the private information contained in our telematics data set. Hence, this model is identical to the model of Chiappori and Salanié (2000), equations (2) and (3). The result shows that we fail to reject the null hypothesis of zero correlation which is consistent with the results of most empirical studies in automobile insurance, see e.g. Chiappori and Salanié (2000) and Dionne et al. (2001). As discussed in Chiappori et al. (2006) and

[5] We also include the interaction term between the number of car rides and the distance driven in both equations to account for the potential interacting effect.

Finkelstein and Poterba (2006), we cannot draw unambiguous conclusions from failing to reject the null hypothesis about the existence and relevance of asymmetric information.

Table 4: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK

	without private information	with private information
ρ	0.0689176	0.054704
	(0.3000)	(0.4173)
β_2		0.0134184
		(0.497)
γ_2		0.0190907
		(0.416)
β_3		0.0009121***
		(0.009)
γ_3		0.0014411***
		(0.000)
N	1849	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

We now use the information contained in our telematics data set and reflected by the two indices *AvgSpeeding* and *#Rides* which is not observed by the insurance company. The second column of Table 4 reports the result of the extended bivariate probit model, equations (6) and (7). The results show that both coefficients β_3 and γ_3 related to the number of car rides are positive and statistically significant. The null hypotheses $\beta_3 = 0$ and $\gamma_3 = 0$ are both rejected at a significance level of less than 1%. This is clear evidence that the number of car rides is positively related to the level of first-party insurance coverage and it is positively related to a downgrade in the Bonus-Malus class. The results on the coefficients β_2 and γ_2 show that average speeding is neither related to the choice of first-party coverage nor to a downgrade in the Bonus-Malus class. Not surprisingly, when adding the private information to the model the residual correlation coefficient ρ remains to be not statistically different from zero.

The relevance of the number of car rides for contract choice and its relation to risk emphasize that rejecting the null hypothesis of zero correlation does not necessarily imply the absence of relevant private information. Moreover, the results that both coefficients β_3 and γ_3 are positive and statistically significant are consistent with the predictions of adverse selection and moral hazard based on private information about driving behavior.

4.2 Liability Insurance Coverage, Driving Behavior, and Risk

In Table 5, we present the results for liability insurance coverage, i.e., the dependent variable *Coverage* in equation (6) is set to 1 for an upper limit of liability coverage equal to € 15m and it is set to 0 if the upper limit is € 10m. The first column of Table 5 reports the correlation coefficient of the error terms ϵ and η when we do not include the private information contained in the telematics data set. The result is equivalent to the result for first-party insurance coverage in Section 4.1: we fail to reject the null hypothesis of zero correlation.

Table 5: LIABILITY COVERAGE, DRIVING BEHAVIOR, AND RISK

	without private information	with private information
ρ	-0.0167378	-0.0072373
	(0.7771)	(0.9036)
β_2		-0.0300407*
		(0.077)
γ_2		0.01932
		(0.410)
β_3		-0.0005385*
		(0.064)
γ_3		0.0014354***
		(0.000)
N	1849	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Next, we add the two indices *AvgSpeeding* and *#Rides* to the bivariate probit model. As with first-party coverage, the results in the second column of Table 5 show that private information contained in the two indices is relevant for the level of liability coverage and is related risk. Consistent with the result in Section 4.1, the relation between the number of car rides and a downgrade in the Bonus-Malus class, as reflected by the coefficient γ_3, is positive and statistically significant. Contrary to the results with first-party coverage, however, the number of car rides is negatively and statistically significantly related to the level of liability coverage. This result is opposite to the predictions of adverse selection or moral hazard and could be explained by selection based on heterogeneous, hidden degrees of risk aversion linked with hidden action (de Meza and Webb, 2001). Policyholders who are more risk-averse purchase more liability coverage, undertake fewer car rides, and are thereby of lower risk.

The result on the coefficient β_2 is consistent with this preference-based selection. β_2 measures the relation between average speeding above legal speed limits and the level of liability coverage in equation (6) and is negative and statistically significant. Policyholders who purchase more liability coverage speed on average less above legal speed limits which both could arise from a higher degree of risk aversion. Interestingly, speeding is not related to a downgrade in the Bonus-Malus class - see the result on coefficient γ_2. This could result from the fact that we underestimate speeding by applying countrywide legal speed limits per road type. In particular, we might underestimate the effect of speeding at street areas which are prone to accidents. Speed limits in these areas are likely to be below the countrywide speed limits. Again, when adding the private information to the model the residual correlation coefficient ρ remains to be not statistically different from zero.

4.3 Results Conditional on Liability Coverage

Motivated by the potential preference-based explanation of the relation between driving behavior and liability choice in Section 4.2, we split our sample according to the level of liability coverage and examine the relation between first-party coverage, driving behavior, and risk separately. The results are presented in Table 6.

Table 6: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK CONDITIONAL ON LIABILITY COVERAGE

	without private information		with private information	
liability coverage	0	1	0	1
ρ	0.0112125 (0.8831)	0.2700989* (0.0537)	-0.0057376 (0.9407)	0.2493633* (0.0859)
β_2			0.0375661 (0.110)	-0.0452905 (0.227)
γ_2			0.0268391 (0.314)	-0.0073328 (0.889)
β_3			0.0009621** (0.018)	0.0009051 (0.205)
γ_3			0.0014503*** (0.001)	0.0016532** (0.044)
N	1293	556	1293	556

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively;
p values are stated in parentheses.

We do find different results for the two groups of policyholders. The first two columns show the results of the bivariate probit model without including the private information of the telematics data set. For the subsample of policyholders with low liability coverage (first column), the correlation coefficient between the level of first-party coverage and a downgrade in the Bonus-Malus class is not significantly different from zero. This result is identical to the one for the entire sample (see Section 4.1). However, for the subsample of policyholders with high liability coverage (second column), we do find a positive and statistically significant correlation coefficient between first-party insurance coverage and a downgrade in the Bonus-Malus class. This correlation could result from an overlaying risk-based and/or the same preference-based selection. The risk-based selection would originate from some additional private information on risk characteristics which overlays the preference selection mentioned above and has a stronger effect on first-party coverage than it has on liability coverage. Both adverse selection and/or moral hazard could explain the residual positive correlation. But, the results could as well be explained solely by the preference-based selection. Individuals who are more risk-averse value insurance coverage more but the effect of risk aversion on the value of risk control is ambiguous (see Ehrlich and Becker, 1972; Dionne and Eeckhoudt, 1985; Jullien et al., 1999). Moreover, if insurance coverage and risk control are substitutes, then a higher level of insurance coverage might reduce the willingness to invest in risk control. Depending on the setting, more risk-averse individuals might as well purchase more insurance coverage but invest less in risk control and thereby be of higher risk.[6]

Liability risk differs from first-party risk in the size of potential financial claims. While the potential financial loss caused by first-party risk is limited by the value of the car, the size of liability claims easily reaches millions of Euros. For risks with potentially severe claims, the substitution effect between between insurance coverage and investment in risk control might be rather small for more risk-averse individuals. Hence, they might purchase more insurance coverage and invest more in risk control. However, for risks with smaller, limited claims, the substitution effect might be larger and more risk-averse individuals might purchase more insurance coverage but invest less in risk control. This could explain both the negative relation between the level of liability coverage and average speeding and the positive relation between the level of first-party coverage and risk for those policyholders who purchase the higher level of liability coverage.

When including the two driving indices *AvgSpeeding* and *#Rides* to the bivariate probit model (columns three and four), the results of the subsample of policyholders with low liability coverage

[6] Jullien et al. (2007) develop a principal-agent model with hidden degree of risk aversion and show that, depending on the parameters, the correlation between insurance coverage and risk can be positive, negative, or zero. Cohen and Einav (2007) present a structural model which accounts for unobserved heterogeneity in both risk and risk aversion. By using a large data set of an Israeli insurance company they find a strong positive correlation between unobserved risk aversion and unobserved risk.

do again not differ from the results of the entire sample: the number of car rides is positively related to both the level of first-party insurance coverage and a downgrade in the Bonus-Malus class. For the subsample of policyholders with high liability coverage, the number of car rides is also positively related to a downgrade in the Bonus-Malus class. However, for the same subsample, the number of car rides is not significantly related to the level of first-party insurance coverage. It is not surprising that the correlation coefficient of the error terms for this subsample is lower when adding private information to the bivariate probit model. It is, however, interesting that it remains positive and statistically significant which suggests residual private information that is not captured by the two driving behavior indices.

5 Robustness and Extensions

Interaction terms

In addition to the interaction terms between the number of car rides and the distance driven, we also include the following interaction terms to account for potential interacting effects of independent variables: $kW * urban$, $kW * male$, $kW * age$, $age * BM$, $year\,of\,construction * kW$, $distance * urban$, $distance * male$, $kW * value\,of\,car$, and $male * value\,of\,car$. Table 7 and 8 in the Appendix show that our results are robust to introducing those interaction terms.

Driving Behavior

Average Speeding

We define the index $AvgSpeeding$ in equation (1) as the speed exceedance above the legal speed limit averaged over the entire number of data points recorded. Since driving behavior might be rather related to the kinetic energy of the car than to the speed, we run the econometric model with squared speed in the index, i.e., with

$$AvgSpeeding^2 = \frac{\sum_j \sum_{i \in \Delta_n} (v_{ij} - u_j)^2}{n}.$$

Table 9 and 10 in the Appendix show that our results are robust with one exception. The negative relation between average speeding and the level of liability coverage which is significant with a p-value of 0.077 for the index $AvgSpeeding$, is no longer significant at the 10% significance level for the index $AvgSpeeding^2$.

Policyholders might also differ in their conditional speeding behavior. Some drivers might constantly drive slightly above the legal speed limit whereas some others might speed rarely but then to a much greater extent. To differentiate this driving behavior, we construct the index $AvgSpeeding$ by averaging the speed exceedances over the number data points at which the driver exceeds the speed limit. The results in Table 11 and 12 show again that our results are robust with the same exception as above. Average speeding above legal speed limits conditional on exceeding the speed limit is not significantly related to the level of liability coverage.

Weekend and Night Driving

We analyze whether differences in the extent to which policyholders drive on weekends or at night explain variation in the level of insurance coverage or risk. For each policyholder, we derive the distance driven on weekends or at night relative to the total distance driven, $\%DistWE$ and $\%DistNight$. We define a weekend drive as a car ride that starts between Saturday 0:00 am and Sunday midnight. A night drive is defined as a car ride that starts between 10:00pm and 6:00 am. We then estimate the following bivariate probit model

$$\begin{aligned} Coverage &= 1(X\beta_1 + \beta_2 AvgSpeeding + \beta_3 \#Rides + \beta_4 \%DistWE + \beta_5 \%DistNight + \varepsilon > 0) \\ Risk &= 1(X\gamma_1 + \gamma_2 AvgSpeeding + \gamma_3 \#Rides + \beta_4 \%DistWE + \beta_5 \%DistNight + \eta > 0). \end{aligned}$$

Table 13 and 14 report the results. First, we observe that our results about the effects of average speeding and number of car rides on contract choice and their relation to risk are robust to the introduction of the two additional indices. Second, neither the percentage distance driven on weekends nor at night are related to the choice of insurance coverage. However, both indices are related to a downgrade in the Bonus-Malus class. While the percentage distance driven at night is negatively related to a downgrade in the Bonus-Malus class, the percentage distance driven on weekends is positively related to it.

We also construct the equivalent indices based on the number of car rides on weekends or at night relative to the entire number of car rides, $\%\#RidesWE$ and $\%\#RidesNight$. The results in Table 15 and 16 show again that the effects of average speeding and the number of car rides are robust to adding these two indices. The percentage number of night rides is also negatively related to a downgrade in the Bonus-Malus class. However, the percentage number of rides during the weekend shows no significant relation to a downgrade in the Bonus-Malus class.

The negative relation between car rides at night and downgrades in the Bonus-Malus class is consistent with the empirical evidence that traffic density is positively related to accident rates (Edlin and Karaca-Mandic, 2006). Our result shows that the effect of lower traffic density at night outweighs the potentially counteracting effect that driving in the dark or while being more tired increases the accident rate.

References

[1] Arrow, K.J., 1963, Uncertainty and the Welfare Economics of Medical Care, *American Economic Review* 53(5), 941-973

[2] Cawley, J., and T. Philipson, 1999, An Empirical Examination of Information Barriers to Trade in Insurance, *American Economic Review* 89(4), 827-846

[3] Chiappori, P.-A., B. Jullien, B. Salanié, and F. Salanié, 2006, Asymmetric Information in Insurance: General Testable Implications, *RAND Journal of Economics* 37(4), 783-798

[4] Chiappori, P.-A., and B. Salanié, 2000, Testing for Asymmetric Information in Insurance Markets, *Journal of Political Economy* 108(1), 56-78

[5] Cohen, A., 2005, Asymmetric Information and Learning in the Automobile Insurance Market, *Review of Economics and Statistics* 87(2), 197-207

[6] Cohen, A., and L. Einav, 2007, Estimating Risk Preferences from Deductible Choice, *American Economic Review* 97(3), 745-788

[7] Cohen, A., and P. Siegelman, 2010, Testing for Adverse Selection in Insurance Markets, *Journal of Risk and Insurance* 77(1), 39-84

[8] Cutler, D.M., and S.J. Reber, 1998, Paying for Health Insurance: The Trade-Off Between Competition and Adverse Selection, *Quarterly Journal of Economics* 113(2), 433-466

[9] Cutler, D.M., and R.J. Zeckhauser, 1998, Adverse Selection in Health Insurance, *Forum for Health Economics and Policy*: Vol. 1: (Frontiers in Health Policy Research), Article 2. http://www.bepress.com/fhep/1/2

[10] De Meza, D., and D.C. Webb, 2001, Advantageous Selection in Insurance Markets, *RAND Journal of Economics* 32(2), 249-262

[11] Dionne, G., and L. Eeckhoudt, 1985, Self-Insurance, Self-Protection and Increased Risk Aversion, *Economics Letters* 17(1-2), 39-42

[12] Dionne, G., C. Gouriéroux, and C. Vanasse, 2001, Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment, *Journal of Political Economy* 109(2), 444-453

[13] Edlin, A.S., and P. Karadac-Mandic, 2006, The Accident Externality from Driving, *Journal of Political Economy* 114(5), 931-955

[14] Ehrlich, I, and G. Becker, 1972, Market Insurance, Self-Insurance, and Self-Protection, *Journal of Political Economy* 80(4), 623-648

[15] Fang, H., M.P. Keane, and D. Silverman, 2008, Sources of Advantageous Selection: Evidence From the Medigap Insurance Market, *Journal of Political Economy* 116(2), 303-350

[16] Finkelstein, A., and K. McGarry, 2006, Multiple Dimensions of Private Information: Evidence From the Long-Term Care Insurance Market, *American Economic Review* 96(4), 938-958

[17] Finkelstein, A., and J. Poterba, 2004, Adverse Selection in Insurance Markets: Policyholder Evidence From the U.K. Annuity Market, *Journal of Political Economy* 112(1), 183-208

[18] Finkelstein, A., and J. Poterba, 2006, Testing for Adverse Selection with Unused Observables, NBER Working Paper No. 12112

[19] Gan, L., M.D. Hurd, and D.L. McFadden, 2005, Individual Subjective Survival Curves, in D. Wise (ed.), *Analysis in the Economics of Aging*, Chicago: University of Chicago Press, 377-411

[20] Hurd, M.D., 1999, Anchoring and Acquiescence Bias in Measuring Assets in Household Surveys, *Journal of Risk and Uncertainty* 19(1-3), 111-136

[21] Hurd, M.D., D.L. McFadden, H. Chand, L. Gan, A., and M. Roberts, 1998, Consumption and Saving Balances of the Elderly: Experimental Evidence on Survey Response Bias, in D. Wise (ed.), *Topics in the Economics of Aging*, Chicago: University of Chicago Press, 353-87

[22] Harris, M., and A. Raviv, 1978, Some Results on Incentive Contracts With Applications to Education and Employment, Health Insurance, and Law Enforcement, *American Economic Review* 68(1), 20-30

[23] Holmstrom, B., 1979, Moral Hazard and Observability, *Bell Journal of Economics* 10(1), 74-91

[24] Jullien, B., S. Salanié, and F. Salanié, 1999, Should More Risk-Averse Agents Exert More Effort?, *The Geneva Papers on Risk and Insurance Theory* 24(1), 19-28

[25] Jullien, B., S. Salanié, and F. Salanié, 2007, Screening Risk-Averse Agents Under Moral Hazard: Single-Crossing and the CARA Case, *Economic Theory* 30(1), 151-169

[26] Pauly, M.V., 1974, Overinsurance and Public Provision of Insurance: The Role of Moral Hazard and Adverse Selection, *Quarterly Journal of Economics* 88(1), 44-62

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

[28] Rothschild, M., and J. Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics* 90(4), 629-649

[29] Saito, K., 2006, Testing for Asymmetric Information in the Automobile Insurance Market Under Rate Regulation, *Journal of Risk and Insurance* 73(2), 335-356

[30] Shavell, S., 1979, On Moral Hazard and Insurance, *Quarterly Journal of Economics* 93(4), 541-562

6 Appendix

Interaction Terms

Table 7: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH INTERACTION TERMS

	without private information	with private information
ρ	0.0685594 (0.3124)	0.0552097 (0.4228)
β_2		0.0136504 (0.494)
γ_2		0.0218825 (0.357)
β_3		0.0008433** (0.016)
γ_3		0.0014868*** (0.000)
N	1849	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Table 8: LIABILITY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH INTERACTION TERMS

	without private information	with private information
ρ	-0.0194319 (0.7458)	-0.0082439 (0.8919)
β_2		-0.031542* (0.067)
γ_2		0.0220499 (0.353)
β_3		-0.0004891* (0.092)
γ_3		0.0014796*** (0.000)
N	1849	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Squared Speeding

Table 9: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH $AvgSpeeding^2$

	with private information
ρ	0.0555587
	(0.4101)
β_2	2.63e-07
	(0.832)
γ_2	1.46e-06
	(0.180)
β_3	0.0007993**
	(0.021)
γ_3	0.0014525***
	(0.000)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Table 10: LIABILITY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH $AvgSpeeding^2$

	with private information
ρ	-0.0066529
	(0.9113)
β_2	-1.34e-06
	(0.186)
γ_2	1.46e-06
	(0.178)
β_3	-0.0005115*
	(0.078)
γ_3	0.0014468***
	(0.000)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Average Speeding Conditional on Exceeding Speed Limit

Table 11: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH *AvgSpeeding* CONDITIONAL ON SPEEDING

	with private information
ρ	0.055178
	(0.4135)
β_2	0.0051087
	(0.596)
γ_2	0.01432
	(0.220)
β_3	0.0007988**
	(0.020)
γ_3	0.0013974***
	(0.000)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Table 12: LIABILITY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH *AvgSpeeding* CONDITIONAL ON SPEEDING

	with private information
ρ	-0.0074615
	(0.9005)
β_2	-0.0103952
	(0.201)
γ_2	0.0143549
	(0.219)
β_3	-0.0004738*
	(0.099)
γ_3	0.001392***
	(0.000)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Percentage Distance Driven on Weekends and at Night

Table 13: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH $\%DistWE$ and $\%DistNight$

	with private information
ρ	0.0606171 (0.3775)
β_2	0.0152704 (0.445)
γ_2	0.0238029 (0.323)
β_3	0.0008603** (0.016)
γ_3	0.0015249*** (0.000)
β_4	-0.3838318 (0.239)
γ_4	0.6607085* (0.086)
β_5	-0.0527865 (0.932)
γ_5	-1.80776* (0.068)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Table 14: LIABILITY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH $\%DistWE$ and $\%DistNight$

	with private information
ρ	-0.0166465
	(0.7843)
β_2	-0.0310278*
	(0.068)
γ_2	0.0242685
	(0.313)
β_3	-0.0005289*
	(0.075)
γ_3	0.0015222***
	(0.000)
β_4	0.064373
	(0.815)
γ_4	0.6609624*
	(0.086)
β_5	0.1431925
	(0.783)
γ_5	-1.799466*
	(0.069)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Percentage Number of Car Rides Driven on Weekends and at Night

Table 15: FIRST-PARTY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH $\%\#RidesWE$ and $\%\#RidesNight$

	with private information
ρ	0.0593909 (0.3848)
β_2	0.0159034 (0.423)
γ_2	0.0212249 (0.377)
β_3	0.0008464** (0.018)
γ_3	0.0015443*** (0.000)
β_4	-0.4587002 (0.178)
γ_4	0.6374247 (0.102)
β_5	-0.2345626 (0.724)
γ_5	-1.969283** (0.042)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.

Table 16: LIABILITY COVERAGE, DRIVING BEHAVIOR, AND RISK WITH $\%\#RidesWE$ and $\%\#RidesNight$

	with private information
ρ	-0.0116968
	(0.8472)
β_2	-0.0310873*
	(0.067)
γ_2	0.0216708
	(0.367)
β_3	-0.0005474*
	(0.065)
γ_3	0.0015404***
	(0.000)
β_4	0.0977193
	(0.705)
γ_4	0.6392511
	(0.101)
β_5	0.3736662
	(0.504)
γ_5	-1.9616**
	(0.043)
N	1849

Notes: significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively; p values are stated in parentheses.