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

Alexander Muermann† Daniela Straka‡

August 2011

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 direct evidence that private information about driving behavior has significant and counteracting effects on the choice of third-party liability and first-party insurance coverage. Yet, tests based on the residual correlation between the level of insurance coverage and risk spuriously allude to different interpretations. These results suggest overlapping and offsetting effects of private information based on risk preferences and driving behavior.

*We are grateful to Randy Dumm, Martin Eling, Alois Geyer, Robert Hauswald, Robert Kremslehner, Christian Laux, Raphael Markellos, Daniel Sturm, and seminar participants at the Risk Theory Society meeting, the ARIA meeting, the European Winter Finance Summit, and the Annual Congress of the German Insurance Science Association for their helpful comments and suggestions.

†Department of Finance, Accounting and Statistics,Vienna University of Economics and Business, and VGSF (Vienna Graduate School of Finance), Heiligenstaedter Str. 46-48, A - 1190 Vienna, AUSTRIA, alexander.muermann@wu.ac.at

‡Department of Finance, Accounting and Statistics, Vienna University of Economics and Business, Heiligenstaedter Str. 46-48, A - 1190 Vienna, AUSTRIA, 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 telematic 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 telematic 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 telematic 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 the following aspects of driving behavior to be significantly linked to contract choice and a subsequent downgrade in the Bonus-Malus class: average speeding above legal speed limits, the number of car rides a policyholder undertakes, and the percentage distance driven on weekends and at night. Average speeding above legal speed limits and the number of car rides are both negatively related to the level of third-party liability coverage. In contrast, the number of car rides and the percentage distance driven at night are positively related to the level of first-party insurance coverage. Last, the number of car rides and the percentage distance driven on weekends are 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

[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 telematic company.

class.

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 between the level of first-party insurance coverage and a subsequent downgrade in the Bonus-Malus class. 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 direct empirical evidence that driving behavior is related to changes in the Bonus-Malus 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 first-party insurance coverage and ex-post realizations of risk. A similar problem in interpreting the residual correlation arises with liability insurance. In this case, we find a statistically significant positive residual correlation between the level of liability coverage. This points to adverse selection and/or incentive effects which are opposite to the preference-based selection effect in liability coverage discussed above. These joint findings support our interpretation of overlapping and counteracting effects of private information about risk type and preferences.

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; McCarthy and Mitchell, 2010). 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; McCarthy and Mitchell, 2010) 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. Chiappori 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 telematic 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

[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.

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 telematic data and insurance data sets and their summary 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 check the robustness of our results. We conclude in Section 6.

2 Data

The insurance company offers a pay-as-you-drive insurance contract in addition to its traditional car insurance contract. Cars insured under this contract are equipped with a telematic 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 factor of 0.8 for pricing the contract. Furthermore, policyholders get a discount on the premium of full comprehensive insurance coverage. In addition to the pay-as-you-drive feature, the telematic 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 telematic device is that stolen cars can be tracked via GPS. Policyholders have to pay a one-time fee for the installation of the telematic device and a monthly fee for the safety services.

2.1 Telematic Data

An independent telematic 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.[3]

[3]Those are all the pay-as-you-drive contracts the insurer has in his portfolio on February 1st, 2009.

For our analysis we restrict our data set to completed car rides, i.e. for which engine switch-on and switch-off were both recorded. 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 summary statistics of the telematic data.

Table 1: SUMMARY 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)
σ speed in km/h (mph)	18.99	22.56	24	35.67
	(11.79)	(14.01)	(14.91)	(22.16)
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 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 telematic 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 comprises all the information used for pricing of the policies in February 2009. Additionally, we have an update of the contract data set for February 2010. We restrict the telematic 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 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.[4]

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).[5]

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 provides the summary insurance statistics.

Table 2: SUMMARY 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.

[4] A downgrade in the Bonus-Malus class is triggered by the submission of at least one liability claim during the year.

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

3 Empirical Approach

3.1 Driving Behavior

We investigate four types of individual driving behavior utilizing the information contained in our telematic data set: average speeding above legal speed limits, the number of car rides, the percentage distance driven on weekends, and the percentage driven at night. The speeding index is given by

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

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.[6]

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. For the other two indices, we derive the distance driven on weekends and at night relative to the total distance driven per policyholder, $\%DistWE$ and $\%DistNight$. For the distance driven on weekends, we use all data points recorded between Saturday 0:00 am and Sunday midnight. For the distance driven at night, we use all data points recorded between sunset and sunrise, using the monthly average as a proxy for both.

We derive all four indices for each car in our data set. The summary statistics for these indices are given in Table 3.

Table 3: SUMMARY STATISTICS INDICES

	first-party coverage		liability coverage		$\triangle BM$ class		
	0	1	0	1	0	1	total
Mean AvgSpeeding	3.05	3.19	3.2	3.02	3.15	3.16	3.15
Mean #Rides	217	255	246	236	238	293	242
Mean %DistWE	0.286	0.26	0.268	0.269	0.266	0.288	0.268
Mean %DistNight	0.083	0.083	0.084	0.081	0.085	0.062	0.083
N	562	1254	1293	556	1708	141	1849

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.

[6] 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.

3.2 Econometric Model

We test for the effect of private information on contract choice and its relation to risk by extending the econometric model suggested by Finkelstein and Poterba (2006). This model is based on the econometric model of Chiappori and Salanié (2000), 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

$$Coverage = 1(X\beta + \varepsilon_1 > 0) \tag{2}$$

$$Risk = 1(X\gamma + \varepsilon_2 > 0) \tag{3}$$

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_1 and ε_2 is zero. 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)

$$Coverage = 1(X\beta_1 + Y\beta_2 + \varepsilon_1 > 0) \tag{4}$$

$$Risk = 1(X\gamma_1 + Y\gamma_2 + \varepsilon_2 > 0) \tag{5}$$

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 provides evidence 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.

Unlike in Chiappori and Salanié (2000) and Finkelstein and Poterba (2006), policyholders in our data set simultaneously choose the level of coverage along two dimensions, first-party and liability coverage. To take into account potential interaction between these two choices, we apply a trivariate probit model. This model consists of three probit regressions based on the Geweke-Hajivassiliou-Keane (GHK) smooth recursive simulator. Interpretation of the results of this trivariate probit model is analogous to the interpretation of the bivariate probit model. We define the dependent variables of the three probits as follows. For liability coverage, we set $CovLiab = 1$ if the upper limit is € 15m and $CovLiab = 0$ if the upper limit is € 10m. For first-party coverage, we set $CovFP = 1$ if the contract covers at-fault losses (full comprehensive insurance) and $CovFP = 0$ otherwise. The dependent variable $\triangle BM$ is set to 1 if the Bonus-Malus level decreased within the subsequent 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.

Y is the set of the four indices $AvgSpeeding$, $\#Rides$, $\%DistWE$ and $\%DistNight$ that characterize driving behavior and are constructed from the telematic data set (see Section 3.1). This information is not observable by the insurance company. We thus apply the following trivariate probit model

$$CovLiab = 1(X\beta_1 + Y\beta_2 + \varepsilon_1 > 0) \tag{6}$$

$$CovFP = 1(X\gamma_1 + Y\gamma_2 + \varepsilon_2 > 0) \tag{7}$$

$$\triangle BM = 1(X\delta_1 + Y\delta_2 + \varepsilon_3 > 0) \tag{8}$$

with

$$Y = (AvgSpeeding, \#Rides, \%DistWE, \%DistNight)$$

and test the null hypothesis of no relevant private information, i.e. for $\beta_2 = 0$, $\gamma_2 = 0$ and/or $\delta_2 = 0$.

We also apply the model of Chiappori and Salanié (2000) by testing for the sign of the correlation coefficients $\rho_{Liab,FP}$, $\rho_{Liab,\triangle BM}$ and $\rho_{FP,\triangle BM}$ of each pair of residual error terms ε_1, ε_2 and ε_3 both excluding and including the set of variables $Y = (AvgSpeeding, \#Rides, \%DistWE, \%DistNight)$.

4 Results and Discussion

Table 4 reports the results of the trivariate probit model, equations (6), (7), and (8).

Table 4: COEFFICIENTS OF TRIVARIATE PROBIT MODEL

	$CovLiab$	$CovFP$	$\triangle BM$
$AvgSpeeding$	-0.0111* (0.0059)	0.003 (0.0069)	0.0026 (0.0031)
$\#Rides$	-0.0002* (0.0001)	0.0003** (0.0001)	0.0002*** (0.0001)
$\%DistWE$	0.0307 (0.0951)	-0.1335 (0.1119)	0.0934* (0.0495)
$\%DistNight$	0.1295 (0.1365)	0.3345** (0.1673)	-0.0463 (0.0802)
kW	0.0015** (0.0008)	-0.0005 (0.0009)	0.0003 (0.0005)
years since construction	0.0116*** (0.0031)	0.1003*** (0.0049)	-0.0073*** (0.0014)
value of car in €	1.41e-06 (2.23e-06)	-1.04e-06 (2.69e-06)	-2.75e-07 (1.34e-06)
urban	-0.0509** (0.0224)	0.0733** (0.0266)	0.0231** (0.0126)
male	0.0053 (0.0236)	-0.178 (0.0281)	0.0016 (0.0123)
Bonus-Malus class	-0.0342 (0.0766)	-0.3012*** (0.0884)	0.043* (0.0384)
age of policyholder	0.0014* (0.0008)	-0.0003 (0.001)	0.0005 (0.0004)
total distance driven	-2.23e-08 (2.08e-08)	7.49e-08*** (2.67e-08)	1.82e-08 (1.08e-08)
Pseudo-R^2	0.0160	0.3532	0.0508
N	1849	1849	1849

Notes: Reported coefficients are marginal effects;
significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively;
heteroscedastic robust standard errors are stated in parentheses.

The coefficients of the four driving indices $AvgSpeeding$, $\#Rides$, $\%DistWE$ and $\%DistNight$ are reported in the first four rows for each of the three probit regressions. In the remaining rows, we report the coefficients of the insurance company's risk classifying variables, the Pseudo-R^2, and the number of observations. The first column reports the coefficients of the liability coverage equation (6), the second column of the first-party coverage equation (7), and the third column of the downgrade in the Bonus-Malus class equation (8). For interpreting the coefficients we only report marginal effects. Both sign and statistical significance of coefficients are identical when estimating the trivariate probit model simultaneously. In our following discussion we focus on the effects of private information contained in the four driving indices.

The results in the first column show that both average speeding and the number of car rides are negatively related to the level of liability coverage. More precisely, speeding on average one km/h more above legal speed limits is related, ceteris paribus, to a 1.11% decrease in the probability of choosing the high liability coverage option. Analogously, undertaking one car ride more in the 3 month observation period is related to a 0.02% decrease in the probability of choosing the high liability coverage option. In the third column, the results show that the number of car rides is a highly statistically significant risk factor. An additional car ride in the 3 month observation period is related to a 0.02% increase in the probability of a subsequent downgrade in the Bonus-Malus class. Note that we do control for the total distance driven. Interestingly, speeding is not related to a downgrade in the Bonus-Malus class. 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 accident since speed limits in these areas are likely to be below the countrywide speed limits. These results on liability coverage and downgrade in the Bonus-Malus class are opposite to the predictions of adverse selection and moral hazard. They could be explained by selection based on heterogeneous, hidden degrees of risk aversion linked with hidden action (de Meza and Webb, 2001) or overconfidence. Policyholders who are less risk-averse or more overconfident purchase less liability coverage, speed on average more, undertake more car rides, and are thereby of higher risk.

In contrast, the results on first-party insurance coverage and downgrade in the Bonus-Malus class are consistent with the predictions of adverse selection and moral hazard. The results in the second column show that the number of car rides is positively related to the level of first-party coverage. Policyholders who undertake more car rides are of higher risk and more likely to purchase full comprehensive insurance coverage. Specifically, undertaking an additional car ride in the 3 month observation period is associated to a 0.03% increase in the probability of choosing full comprehensive insurance coverage.

Finally, the results show that the percentage driven at night is positively related to the level of first-party coverage and that the percentage driven on weekends is a statistically significant risk factor.

It is interesting that the number of car rides is a highly significant risk factor, controlling for the distance driven. A possible explanation is that the start and the end of a car ride are particularly exposed to accident risk since the driver has to fulfill multiple tasks such as pulling out the car into the passing traffic or parking the car which involves slowing down, potentially looking for a parking spot, and reversing into it. These simultaneous tasks might demand much more attention from the driver and are thus more prone to accidents than the task of driving the car in the traffic.

In summary, the results of the trivariate probit model show that there exists private information contained in the four driving indices that is relevant for contract choice and risk measured as a subsequent downgrade in the Bonus-Malus class. Furthermore, the effects related to third-party liability coverage are opposite to the effects related to first-party coverage. The results suggest a negative association between the level of liability coverage and risk, while they suggest a positive association between the level of first-party coverage and risk. These opposing correlation signs could result from an overlay of risk-based and preference-based selection. The risk-based selection originates from private information on risk characteristics which overlays the selection based on preferences such as risk aversion. Since the potential severity of liability claims is much higher than the one of first-party claims, the preference-based selection might have a relatively stronger effect on liability coverage than it has on first-party coverage. Differences in the degrees of risk aversion might be a much more important factor when facing claims in millions of € than when facing a loss that is restricted by the value of the car. This would explain the opposite effects on liability and on first-party coverage.[7]

Table 5 reports the correlation coefficients of the residual error terms.

[7] While it is true that individuals who are more risk-averse value insurance coverage more, 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. 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 which strengthens the positive correlation property.

Table 5: CORRELATIONS OF RESIDUAL ERROR TERMS

	without private information	with private information
$\rho_{Liab,FP}$	0.113** (0.0106)	0.128*** (0.0041)
$\rho_{Liab,\triangle BM}$	0.061* (0.0822)	0.06* (0.0944)
$\rho_{FP,\triangle BM}$	-0.015 (0.7311)	0.000 (0.9967)
N	1849	1849

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

For comparison, we first test for the positive correlation property between insurance coverage and risk as if we did not have access to the additional private information contained in Y. The first column reports the correlation coefficients when excluding the four driving indices from the trivariate probit model. This model is thus a trivariate version of the model of Chiappori and Salanié (2000), equations (2) and (3). The results show that we fail to reject the null hypothesis of zero correlation between first-party coverage and a downgrade in the Bonus-Malus class, $\rho_{FP,\triangle BM} = 0$, 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 Finkelstein and Poterba (2006), we cannot draw unambiguous conclusions from failing to reject the null hypothesis about the existence and relevance of asymmetric information. And this is exactly confirmed by our results. Although we fail to reject the null hypothesis of zero correlation between first-party coverage and a downgrade in the Bonus-Malus class, we do find that private information, in particular the number of car rides, is relevant for the level of first-party insurance coverage and for a downgrade in the Bonus-Malus class (see Table 4).

A similar conclusion must be drawn about interpreting the statistically significant positive correlation of the residual error terms between liability coverage and a downgrade in the Bonus-Malus class, $\rho_{Liab,\triangle BM}$. The positive sign of the correlation coefficient is new to the literature which has focused on first-party coverage. This could be interpreted as solely arising from adverse selection and/or incentive effects. However, the negative relation of both average speeding and number of car rides to the level of liability insurance coverage (see Table 4) suggest at least an additional preference-based selection effect which is opposite to the one of adverse selection. The positive correlation coefficient in conjunction with the results on average speeding and the number of car rides is thus another indication of overlaying risk-based and preference-based selection effects.

Last, the correlation between the residual error terms of the liability and first-party coverage equations $\rho_{Liab,FP}$ is highly statistically significant and positive. This is consistent with some private information, such as risk aversion, which explains why policyholders who choose full comprehensive coverage also choose the high liability coverage option.

The second column in Table 5 reports the correlation coefficients between the residual error terms when including the four driving indices in the trivariate probit model. The results do not change. The correlation coefficient $\rho_{FP,\triangle BM}$ between the error terms of first-party coverage and risk remains to be not statistically different from zero. Similarly, the correlation coefficients $\rho_{Liab,FP}$ and $\rho_{Liab,\triangle BM}$ between the error terms of liability and first-party coverage and between liability coverage and risk remain statistically significant and positive.

5 Robustness Checks

Selection Bias

Since the pay-as-you-drive insurance contract is offered for choice, there might be a selection bias if the characteristics of policyholders who choose the pay-as-you-drive insurance contract are correlated with our three dependent variables. To control for this potential selection bias, we employ the two-step Heckman correction method based on an additional data set of randomly selected 2000 cars which are insured under the traditional insurance contract. The policyholders contained in this data set thus decided not to switch to the pay-as-you-drive contract. Data cleaning leaves us with 1987 traditional insurance contracts. Table 6 provides the summary insurance statistics under the traditional insurance contract.

In the first stage, we estimate the following probit regression model

$$selection = 1(X\gamma_1 + \varepsilon > 0)$$

based on the both samples of policyholders, the randomly selected sample of those who chose not to switch and the sample of those who did switch to the pay-as-you-drive insurance contract. *selection* is a binary variable, equal to 1 if the policyholder chose the pay-as-you-drive insurance contract and equal to 0 if the policyholder chose the traditional insurance contract. X consists of all the variables used by the insurance company for pricing the traditional insurance contract. Table 7 reports the results of the first-stage regression. It shows that all the variables are highly

Table 6: SUMMARY STATISTICS TRADITIONAL INSURANCE DATA

	Mean				
	total	none/compr.	full compr.	liab. 10m	liab. 15m
car's characteristics:					
years since construction	5.32	7.18	0.94	5.53	4.83
kW (HP)	75.83	74.41	79.17	75.49	76.62
	(101.69)	(99.79)	(106.17)	(101.23)	(102.75)
value of car in €	22,768	22,615	23,127	22,588	23,188
policyholder's characteristics:					
age in years	54.65	54.82	54.25	54.57	54.83
male	0.33	0.30	0.37	0.32	0.33
urban	0.21	0.19	0.25	0.20	0.21
BM (Bonus-Malus class)	0.46	0.46	0.45	0.46	0.45

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.

significant on the selection of the type of insurance contract. The pay-as-you-drive insurance contract is more likely to be chosen by drivers of cars with more engine power, that are older, and more valuable. Moreover, younger, male policyholders who live in cities with a higher Bonus-Malus class are more likely to chose the pay-as-you-drive contract.

Table 7: RESULTS FOR FIRST-STAGE SELECTION MODEL

	Coefficients.
kW	0.0019*** (0.0006)
years since construction	0.0195*** (0.0018)
value of car in €	3.86e-06** (1.62e-06)
urban	0.2692*** (0.0171)
male	0.0666*** (0.0185)
Bonus-Malus class	1.6062*** (0.1061)
age of policyholder	-0.0066*** (0.0006)
Pseudo-R^2	0.1733
N	3985

Notes: coefficients reported are marginal effects; standard errors are stated in parentheses; significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively;

We then derive the inverse Mill's ratio

$$InvMills = \frac{\phi(\omega\gamma)}{\Phi(\omega\gamma)}$$

where $\phi(\omega\gamma)$ is the standard normal probability density function and $\Phi(\omega\gamma)$ is the standard normal cumulative distribution function of the z-score of each observation estimated in the first stage.

In the second stage, we include the inverse Mill's ratio as an additional variable in the trivariate probit model to control for the potential selection bias. We use the variable *male* as an instrumental variable since it is not significant in the trivariate probit regression. *male* is thus excluded in the second stage. The gender variable in our data set indicates the gender of the person that purchased the insurance contract. This person is thus likely to be the one who decides whether to switch to the pay-as-you-drive insurance contract. However, this person is not necessarily the person who mainly drives the car since the insurance contracts are related to the car and not to the specific driver. In fact, all losses are insured that are caused by any person who drives the car with the approval of the policyholder. This might explain the significance of gender in the selection of the insurance contract but the lack of significance in the second-stage regression.

The second-stage trivariate probit model is thus

$$CovLiab = 1(X\beta_1 + Y\beta_2 + \beta_3 InvMills + \varepsilon_1 > 0) \tag{9}$$

$$CovFP = 1(X\gamma_1 + Y\gamma_2 + \gamma_3 InvMills + \varepsilon_2 > 0) \tag{10}$$

$$\triangle BM = 1(X\delta_1 + Y\delta_2 + \delta_3 InvMills + \varepsilon_3 > 0) \tag{11}$$

Table 8 reports the coefficients of the second-stage trivariate model and shows that our results on the relevance and effects of private information contained in the four driving indices *AvgSpeeding*, *#Rides*, *%DistWE* and *%DistNight* are robust when controlling for a selection-bias. The statistically significant positive coefficient of the inverse Mill's ratio in the first-party coverage probit regression indicates that policyholders who opt for the pay-as-you-drive contract are less likely to choose the full comprehensive insurance contract.

Table 8: COEFFICIENTS OF SECOND-STAGE TRIVARIATE PROBIT MODEL

	$CovLiab$	$CovFP$	$\triangle BM$
$AvgSpeeding$	-0.0111* (0.0059)	0.0034 (0.0072)	0.0026 (0.0031)
$\#Rides$	-0.0002* (0.0001)	0.0002* (0.0001)	0.0002*** (0.0001)
$\%DistWE$	0.0302 (0.0955)	-0.1282 (0.1119)	0.0929* (0.0583)
$\%DistNight$	0.1296 (0.1371)	0.3492** (0.1652)	-0.048 (0.0768)
kW	0.0015* (0.0008)	0.0001 (0.0012)	0.0002 (0.0004)
years since construction	0.0118*** (0.0039)	0.1066*** (0.0074)	-0.008*** (0.0017)
value of car in €	1.35e-06 (2.19e-06)	2.60e-07 (3.21e-06)	-4.49e-07 (1.12e-06)
urban	-0.0489 (0.0424)	0.1587*** (0.0463)	0.0123 (0.019)
Bonus-Malus class	-0.0277 (0.1692)	0.0251 (0.1838)	0.0024 (0.0682)
age of policyholder	0.0013 (0.0013)	-0.0027* (0.0014)	0.0008* (0.0006)
total distance driven	-2.23e-08 (2.23e-08)	7.51e-08*** (2.47e-08)	1.82e-08 (1.12e-08)
inverse Mill's ratio	0.0049 (0.0994)	0.2388** (0.115)	-0.0288 (0.0445)
Pseudo-R^2	0.0159	0.3549	0.0512
N	1849	1849	1849

Notes: coefficients reported are marginal effects;
significance levels are labeled ***, ** and * at 1%, 5% and 10% respectively;
heteroscedastic robust standard errors are stated in parentheses.

Table 9 shows that the results for the correlation coefficients of the residual error terms are also robust with one exception. When adding private information to the model, the correlation coefficient $\rho_{Liab,\triangle BM}$ between liability coverage and downgrade in Bonus-Malus class loses its significance.

Table 9: CORRELATIONS OF RESIDUAL ERROR TERMS

	without private information	with private information
$\rho_{Liab,FP}$	0.12*** (0.006)	0.134*** (0.0027)
$\rho_{Liab,\triangle BM}$	0.059* (0.0876)	0.057 (0.1206)
$\rho_{FP,\triangle BM}$	-0.003 (0.938)	0.009 (0.8478)
N	1849	1849

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

6 Conclusions

We capitalize on having access to detailed data on driving behavior of policyholders in automobile insurance which is inaccessible for the insurance company. By connecting this data to insurance data, we provide direct evidence that driving behavior is relevant for contract choice in first-party and third-party liability insurance as well as for risk. Whereas the number of car rides and average speeding above legal speed limits is negatively related to the level of liability coverage, the number of car rides and the percentage distance driven at night are positively related to the level of first-party insurance coverage. Moreover, the number of car rides and the percentage distance driven on weekends are significant risk factors. These results pulled together suggest the coexistence and interaction of risk-based and preference-based selection effects.

We then test for the residual correlation between insurance coverage and risk which would be the standard test for asymmetric information if we did not have access to the data on driving behavior. The results emphasize that the residual correlation test can be misleading when interpreted in the context of asymmetric information. We fail to reject the hypothesis of zero residual correlation between first-party coverage and risk although the number of car rides is positively related to both first-party insurance coverage and risk. Similarly, we find a significant positive residual correlation between liability coverage and risk although the number of car rides is negatively related to liability coverage but positively related to risk.

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

[14] 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

[15] 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

[16] 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

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

[18] 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

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

[20] 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

[21] 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

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

[23] 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

[24] 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

[25] McCarthy, D., and O.S. Mitchell, 2010, International Adverse Selection in Life Insurance and Annuities, in S. Tuljapurkar, N. Ogawa, and A.H. Gauthier (eds.), *Ageing in Advanced Industrial States: Riding the Age Waves*, Vol. 3, Springer, 119-135

[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