The Private Information in Automobile Insurance Market - Revealed by Driving Record in Telematics

Yinglu Deng[1] Kili C. Wang[2] Tianyang Wang[3] Hao Zheng[4]

[1]Department of Finance, Tsinghua University. Email: dengyl@sem.tsinghua.edu.cn

[2]Department of Insurance, Tamkang University, Research Fellow at the Risk and Insurance Research Center, National Chengchi University. Email: kili@mail.tku.edu.tw

[3]Department of Finance and Real Estate, Colorado State University. Email: Tianyang.wang@business.colostate.edu

[4]Department of Finance, Tsinghua University. Email: zhengh.13@sem.tsinghua.edu.cn

The Private Information in Automobile Insurance Market - Revealed by Driving Record in Telematics

Abstract

By using the telematics information, we find that the driving behaviors recorded by the telematics contain valuable private information for auto insurers. We first confirm the findings from the prior literature that high driving amount and speeding indicate high driving risk. Furthermore, we identify for the first time that frequently rapid acceleration/deceleration is an even more powerful index for high driving risk, and risky driving behaviors indicated by the tendency of driving faster than others is a good index for less risk aversion. Merging this information with insurance data set, we find that the multi-dimension asymmetric information problems, including the adverse selection/moral hazard and the advantageous selection, can be identified through the new approach of "private information test" with the help of the private information we explored. Additionally, our research provides the evidence that individuals tend to react differently when they are merely observed versus when they have the concerns of premium changes in their insurance. This new evidence supports the view that the UBI contract can help incentivize the drivers to adopt safer driving habits and reduce the traffic accident.

Keywords: Asymmetric information, Private Information, Telematics, Big Data, Automobile Insurance

The Private Information in Automobile Insurance Market - Revealed by Driving Record in Telematics

1 Introduction

Collecting the driving behaviors recorded by the telematics, this paper provides valuable private information of the drivers which can help to decode the asymmetric information problems in automobile insurance. The new private information revealed from this research not only help to mitigate the asymmetric information problem confronted by the insurers, it also contributes to the increasingly popular automobile user based insurance (UBI) contract by providing meaningful risk factors for it.

Since the seminal work of Akerlof (1970), asymmetric information has been a core research topic of modern economics (c.f., Bouvard 2014; Roberts 2015). The insurance industry works on the principle of risk and information and the insurance market has been considered as a near ideal context for exploring the prediction of asymmetric information. Rothschild and Stiglitz (1976) first introduce the notion of adverse selection in insurance markets. Since then, voluminous empirical studies have emerged in this burgeoning area of of research (c.f. Meza and Webb 2001; Finkelstein and McGarry 2006). Cohen and Siegelmann (2010) provide a comprehensive review of the empirical literature on asymmetric information in insurance markets.

Most of the literature on asymmetric information in insurance market primarily focus on the adverse selection and moral hazard problems based on insurance data (e.g. insurance coverage and ex post realizations of risk) to reveal the private information of individual's information advantage before purchasing insurance or the individual's behaviors after purchasing insurance. While the early studies mainly examine single dimension of private information (c.f., Harris and Raviv, 1978; Cutler and Reber, 1998), Meza and Webb (2001) point out individual's risk aversion as another possibility of private information. More recent literature has paid closer attention to multiple dimensions of private information.

Finkelstein and McGarry (2006) find that the unobserved preference heterogeneity may offset the positive correlation between the insurance coverage and risk raised by traditional asymmetric information problem, hence the asymmetric information problem could still exist even when there is no significant conditional correlation between the insurance coverage and risk. Therefore, the approach for testing asymmetric information problem confronts revolution, and the new criterion of "private information test" is shown to be more appropriate comparing to the classic criterion of "correlation test". The "private information test" examines whether the asymmetric information problem exists by the individual's private information not held or used by the insurers and is correlated with both insurance coverage and risk occurrence. It is evident that collecting private information in different dimensions become increasingly important.

Valuable private information often hand collected by researchers. Finkelstein and McGarry (2006) collect the measurement of individual beliefs about nursing home use, wealth level, and cautiousness behaviors (e.g., flu shot, health test, wear seat belt, etc.) using the survey of Asset and Health Dynamics Among the Oldest Old (AHEAD). They show that the individual beliefs are private information related to risk type, and the wealth level and cautiousness behaviors are private information related to risk preference. Wang et al. (2009) collect the firm's self-protection behaviors related to fire risk by questionnaires[1], and report that the firm's self-protection behaviors are significantly negatively related to the fire risk and positively related to the willingness of purchasing fire insurance. Similarly, Wang et al. (2011) collect individual's family cancer history by questionnaires, and show that the individual's family cancer history provide valuable information related to individual's risk type in cancer insurance and can help mitigate the adverse selection problem.

Furthermore, Brockett and Golden (2007) suggest that the records of individual's behaviors in different dimensions are correlated because that they are all reflection of individual's biological and psychological characteristics. They find that individual's credit records are good proxy of risk in different kinds of insurance. Bair et al. (2012) use driver's proper maintenance behavior as proxy for low risk drivers. More recently, Huang and Wang (2016) find that the driving risk can also be revealed by individual's vehicle choosing decision. The green car drivers are proved to be low risk drivers.

[1] "Qualified operation of the fire safety equipment" and "self-defense fire organization" are two proxies to indicate firm's self-protection behaviors.

With the ever increasing amount of data being generated and to be assessed, analyzing big data has become a key basis of competition, underpinning new waves of productivity growth and innovation. Telematics Auto Insurance is such a recent innovation by auto insurers that potentially could provide more accurate assessment of risks by more closely observing and examining the private information related to individual's driving behaviors and risk preference. A driver's behavior is monitored directly while the person drives. Driving behaviors are tracked using in-vehicle telecommunication devices (telematics) that are either installed into a special vehicle port or already integrated in original equipment installed by car manufactures. These telematics devices measure a number of elements of interest to underwriters: miles driven; time of day; where the vehicle is driven (GPS); rapid acceleration; hard braking; hard cornering; and air bag deployment. The level of data collected generally reflects the telematics technology employed and the policyholder's willingness to share personal data. By using such private information, insurers can develop the contract of UBI and more closely align premium rates of automobile insurance with driving behaviors.

The extant literature on UBI primarily focuses on the benefit and modification of "pay-as-you-drive" as compared to the traditional automobile insurance. To the best of our knowledge, Kremslehner and Muermann (2016) is the only one literature exploring the private information recorded by telematics and the asymmetric information problems in automobile insurance. They examine the private information related to the coverage choice and claim including average speeding, number of rides, and relative driving distance at night or on weekend. Based on a unique telematics data set, we also collect the private information of driving behaviors, and propose a general test of the presence of asymmetric information in auto insurance industry.

Our work is distinguished from and complements the findings of Kremslehner and Muermann (2016) in two important aspects: 1) By collecting more comprehensive private information including sudden increase/decrease in the driving speed, we are the first research to identify that frequently rapid acceleration/deceleration is an even more powerful index for high driving risk, and risky driving behaviors indicated by the tendency of driving faster than others is a good index for less risk aversion. 2) It is important to note that the drivers in the research sample of Kremslehner and Muermann (2016) are also the insured of the UBI contract. Therefore, the drivers' driving behaviors in Kremslehner and Muermann (2016) are potentially affected by the concerns of triggering the premium changes of their insurance contracts. In comparison, we collect the driving records from the "neutral" drivers who are not the insured of the UBI contract. Naturally, our empirical findings complement Kremslehner and Muermann (2016) and provide a "neutral" comparison with Kremslehner and Muermann (2016) for the insured covered by the UBI contract. In addition, our research provides the evidence that individuals tend to react differently when they are merely observed versus when they have the concerns of price penalty in insurance. For instance, some driving behaviors are restrained so as these related private information to be less influential in Kremslehner and Muermann (2016), but they are much more prominent in our research. This new evidence helps to support the view that UBI contract can incentivize the drivers to adopt safer driving habits and reduce the traffic accident.

We find that higher road usage amount and radical behavior on the road indicate the high risk driver. They relate to both of the insurance coverage and the risk occurrence in the same direction. Some incautious behavior indicates that the driver is less risk aversion. So, it positively relates to risk occurrence and inversely relates insurance coverage. By using the new approach of "private information test" through the private information collected by the telematics, we find that adverse selection/moral hazard exists even when the conditional correlation between coverage and risk is significantly negative. Hence, we also echo to the literature that it is multiple dimension asymmetric information problems existed in the market. Furthermore, they are significant factors in traffic accident risk which can also be applied to UBI products. Our research provides new insights to help: 1) better understand the big data from telematics for auto insurers; 2) better decode the asymmetric information problem confronted by the insurers; 3) provide meaningful risk factors for the increasingly popular UBI; and 4) provide incentive for UBI insured to adopt safety driving habit.

The paper is structured as follows. Section 2 provides background information on the development of UBI contracts. Section 3 provides detailed data description, and how the information of driving behaviors is generated from the records of telematics. Section 4 presents our empirical models and empirical results. Finally, Section 5 summarizes and concludes the paper.

2 Background

The UBI is a growing segment of the automotive insurance market and the emerging contract of UBI provides a possible solution for the asymmetric information problem. The first UBI programs surfaced in the U.S. about a decade ago, when Progressive Insurance Company and General Motors Assurance Company (GMAC) began to offer mileage-linked discounts through combined GPS technology and cellular systems that tracked miles driven. Recent developments in technology have increased the effectiveness and decreased the cost of using telematics, enabling insurers to capture not just how many miles people drive, but how and when they drive. The result has been the growth of several UBI variations, including Pay-As-You-Drive (PAYD), Pay-How-You-Drive (PHYD), Pay-As-You-Go, and Distance-Based Insurance.

According to IHS Automotive (IHS 2016)[2], a leading source of critical information and insight to the global automotive industry, close to 12 million consumers globally subscribed to the automobile UBI contracts in 2015, and they expect that the number of consumer subscribers to UBI is expected to grow to 142 million globally by 2023. America is by far the largest car insurance market in the world, with more than 260 million vehicles in operation and more than 5 million UBI policyholders in 2015. Italy was a distant second, with 3.6 million subscribers out of 36.8 million vehicles in operation in 2015. Italy and the United Kingdom are the most mature UBI markets in Europe, with growing but smaller markets in France, Germany and Spain. Another key market to watch is China. The Chinese government began granting foreign insurance carriers access to the market in 2012. According to IHS Automotive forecasts, the Chinese UBI subscriber volumes will grow from 50,000 in 2015 to over 22 million by 2023 due to sheer size of China's automotive market.

The UBI programs offer many advantages to both insurers and insured. The use of telematics helps insurers monitor driver safety, more accurately estimate accident damages and reduce fraud by enabling them to analyze the driving data (such as hard braking, speed, and time) during an accident. This additional data can also be used by insurers to refine or differentiate UBI products. Moreover, the ancillary safety benefits (e.g., roadside assistance and vehicle theft recovery) offered in conjunction with many telematics-based UBI programs help lower accident rate by improving accident response time, and reduce vehicle theft related costs by allowing for stolen vehicles to be

[2] Auto Tech Report - Usage Based Insurance (UBI) – 2016, IHS Automotive.

tracked and recovered. When UBI is used on fleets, telematics can also help improve the fleets' performance by allowing the fleets to determine the most efficient routes, saving them costs related to personnel, gas, and maintenance. Furthermore, the insurer can also link the insurance premium to the performance of the individual driver when UBI is adopted. Linking insurance premiums more closely to actual individual vehicle allows insurers to more accurately price premiums, and increases affordability for lower-risk drivers. It also gives consumers the ability to control their premium costs by incentivizing them to reduce miles driven and adopt safer driving habits. Fewer miles and safer driving also aid in reducing accidents, congestion, and vehicle emissions, which benefits society. On the other hand, consumers are still unfamiliar with the UBI product and privacy concerns often make consumers wary of sharing data.

Most of the extant literature of automobile UBI highlights the concept of the differences between the traffic risk per mile driven and the traffic risk per vehicle-year, or follow with interest on the differentiated driving risk in different timing within a day and different location (c.f. Litman 2005; Ayuso et al. 2016). For instance, Litman (2005) point out that the traffic accident risk measured by per vehicle-year could be different with the risk measured by per mile driven, and suggest that some low (high) risk insured could come up with higher (lower) risk in the whole year because they drive more (less). Ayuso et al. (2016) find that the urban driving, nighttime driving, and speeding have significantly positive effect on risk. Litman (2012) consider to design a differentiated pay-as-you-drive with driving in urban area get the penalty. Kremslehner and Muermann (2016) and our research focus on exploring private information recorded by telematics and the asymmetric information problems in automobile insurance. Average speeding, number of rides, relative driving distance at night or on weekend provide private information related to the coverage choice and claim (Kremslehner and Muermann 2016). Such private information has multiple and often counteracting effects. In this research, we explore new private information and provide a supplementary research for Kremslehner and Muermann (2016) by investigating in a "neutral" research sample that the drivers are not yet the insured of UBI contract. Our work and Kremslehner and Muermann (2016) together provide an opportunity to compare the reaction of the drivers between merely be observed or further concern the price penalty. Further discussion is in Section 5.

3 Data

Although the formal UBI contract has not yet surfaced in China, confronting the worldwide development of this new trend, some insurance companies in China have tried to collect the driving information from the drivers by encouraging their insured to equip with the in-vehicle telecommunication devices (telematics). They aim at designing the feasible UBI contracts after they get sufficient information from analyzing the information collected by these telematics.

We obtain the driving behaviors data from the telematics company "Yesway" during December 2, 2014 and March 31, 2016. If the drivers buy and install the OnBoard Diagnostics (OBD) telematics device from Yesway, Yesway will provide navigation and other services according to the telematics data (similar to OnStar). Note that Yesway is independent from insurance companies and insurance companies are neither involved in telematics device services nor shared the telematics data. Therefore, the driving records are from the "neutral" drivers who are not the insured of the UBI contract.

In order to merge the driving behavior information with the insurance data set, we obtain insurance data from the Chinese Insurance Information Technology Company (CIITC). CIITC is the insurance industry's big data platform in China. It collects data from all insurance companies, and has the most comprehensive insurance data from China's insurance industry. Equipped with the CIITC data, we matched the vehicle equipped with telematics device from Yesway in Beijing area with the associated insurance information from the top three insurance companies in China (Renbao, Pingan, and Taiping). We only keep the insured whose contract is completed[3] and only use telematics for more than 30 days in order to get stable description about their usage and driving behaviors. We connect the driving behavior records with the information collect from their insurance contracts in our sample period. The Chinese automobile insurance market was regulated with three main kinds of contracts during our sample period. All contracts have the same pricing matrix. In year 2016, there are 8268 of insured equipped with the telematics and sent back their driving behavior messages. We have total of 7785 observations. The variables we collect for this research are defined and listed in Table 1.

[3] The average effective time of contracts is 242.28 days.

[Insert Table 1 Here]

The driving behaviors recorded by the telematics includes: their road usage amount, radical behavior on the road, and cautious behavior. The road usage amount is measured by the number of rides per day. The driver's radical behavior on the road is measured by the frequencies of rapid acceleration and rapid deceleration. The rapid acceleration/rapid deceleration behavior is classified into four categories according to the 25% quantiles of rapid acceleration frequencies and medium of rapid deceleration frequencies. The four categories include: the drivers who are high frequency in rapid acceleration and high frequency in rapid deceleration, the drivers who are high frequency in rapid acceleration and low frequency in rapid deceleration, the drivers who are low frequency in rapid acceleration and high frequency in rapid deceleration, and the drivers who are low frequency in rapid acceleration and low frequency in rapid deceleration. The fourth category is the counter part of the first three categories. We measure the cautious behavior of individual when he drives. The cautious driving behavior of drivers measured here is indicated by keeping the "right speed" on the road. Keeping the "right speed" means to keep on or around average speed. Hence, we measure the cautious driving behavior here by keeping the driving speed above the medium.

The contract we explore is the voluntary liability insurance. The coverage of this contract varies from 50,000 RMB to 2 million RMB. The structure of the insurance amount in our research sample is listed in Table 2. We define the coverage of the contracts over 0.3 million RMB as high coverage contracts.[4]

[Insert Table 2 Here]

The information collected from the contract includes: age, gender, employment status, previous three years of claim record, insured status in comprehensive insurance and the customer loyalty of the insured; the age and the passenger load factor of the insured vehicle; the distribution channel of the contract.[5] The traffic accident risk of each insured is measured by the claim records during

[4]We define the higher versus low coverage contract by the median of coverage.

[5]The age of the insured is classified into 5 categories. The counter part of the age classification variables is the group of insured who are older than 60 years old. The employment status of the insured is classified into three categories. The counter part of government employee and corporate employee is else (this is a standard category used by insurance company). The previous three years of claim records is represented by a bonus malus coefficient. The higher this coefficient means the poorer the insured's claim record is. The insured status is that whether the

the policy period of these contracts. The claim records include not only the claim filed during our research period from the voluntary liability insurance contract, but also include the claim filed during our research period from the compulsory liability insurance contract of the same insured.[6]

As for the driving message sent back from the telematics, the information of driving amount and driving habit are included. In the information of driving amount, we can observe it from the dimension of the use rate (the percentage of driving days during our research period, *rate_use*), as well as the driving frequencies (average number of rides per driving dy, *N_rides*), driving mileage (the average driving mileage per driving day, *avmlg*) and the average time of driving per driving day (*time_drv*).[7] In the information of driving habit, we can observe it from the dimension of average driving speed (the average driving speed per driving day, *avspd*), over speed ratio (measured by the time of speeding over the time of driving, *speeding*), the level of maximum driving speed (*maxspd*), the frequencies of rapid acceleration driving per day (*freq_acc*),[8] the frequencies of rapid deceleration per driving day (*freq_dec*),[9] the ratio of driving at night (measured by the time of driving at night over the time of driving, *ratio_night*).[10] The overall observation related to above information in our research sample is in panel A of Table 3.

[Insert Table 3 Here]

Among the fruitful information related to the driving amount and driving habit, we need to choose some information and define some variables as representatives. The first task is to classify them according to their attributes: *rate_use*, *N_rides*, *avmlg* and *time_drv* are classified as the categories of driving amount; *speeding*, *freq_acc*, *freq_dec* , are classified as the categories of radical driving behavior; *avspd*, *maxspd*, *ratio_night* are classified as other factors related to individual driver's risk type or risk preference.

insured has further purchased the comprehensive insurance contract. The customer loyalty is measured according to whether he/she has continued to insure from the same insurance company. The age of the vehicle is classified into four categories. The counter part of these car age variables is the group of vehicles which are over 9 years. The vehicle's passenger load factor is classified according to whether the vehicle's seats are less than 6 or not. The distribution channel is classified by whether the contract is sold through telephone marketing. These variables which are controlled in our research are chosen according to the rule of underwriting and pricing of the insurance companies.

[6] It is worth to be mentioned that the claim of the same insured recorded in different contracts but happened in the same day is recorded as one time of accident.

[7] The driving frequencies (N_rides), driving mileage (avmlg) and the length of driving time (time_drv) per day are measured by the actual driving days instead of calendar days during our research period.

[8] The "rapid acceleration" is defined by the increment of speed that is more than 10km/h within one second.

[9] The "rapid deceleration" is defined by the reduction of speed that is more than 10km/h within one second.

[10] It is measured by driving time during 8pm—6am over the sum of driving time.

The second task when we choose meaningful or significant factors from each category to define variables for our research, we observe the correlations between them and *claim/cov_high*. The simple correlations are observed from the correlation coefficients which are listed in Table 4. The conditional correlations are observed from their estimated coefficients when they are one by one put in regressions of bivariate probit model together with the other characteristics of the insured and the insured vehicle. And each pair of estimated coefficients of them in each single bivariate probit model is listed in Table 5.

[Insert Table 4 and Table 5 Here]

In the category of driving amount, we find that *N_rides* has the highest and most significant correlation with *claim/ cov_high*, both observed from simple correlation and from conditional correlation. Following Kremslehner and Muermann (2016) we also choose *N_rides* as the representative of driving amount, and control the information of distance of driving by total driving mileage *avmlg*. Note that *avmlg* has significant correlation and conditional correlation with *claim/ cov_high* from Table 4 and Table 5.

In the category of driving habit, we find that both *freq_acc* and *freq_dec* have the highest and most significant simple correlation with *claim/cov_high*. Their conditional correlations are also the most significant. Furthermore, *freq_acc* and *freq_dec* are strongly correlated to each other (the correlation coefficient is 0.4906, and is significantly different from 0). We propose the index (*radical*) to indicate an individual's tendency of radical driving behavior which is related to rapid acceleration or rapid deceleration. The *radical* index is a set of three dummy variables which represent the individual's degree of radical: the tendency of frequently rapid acceleration as well as rapid deceleration (*radical_1*), the tendency of frequently rapid acceleration but not frequently rapid deceleration (*radical_2*), the tendency of frequently rapid deceleration but not frequently rapid acceleration (*radical_3*), and the tendency of neither frequently rapid deceleration nor frequently rapid acceleration (this is the counter group of *radical_1*, *radical_2* and *radical_3*).[11]

As for the other factors, we find that *avspd* and *maxspd* are both significantly correlated with *claim/ cov_high* when we observe from the simple correlation. However, *avspd* seems to indicate

[11] We use 25% quantiles to be the threshold of high versus low frequency of rapid acceleration, and use medium to be the threshold of high versus low frequency of rapid deceleration is the best. We also tried different levels of threshold and our results are robust.

the individual's risk preference because the sign of the correlation is opposite when it correlates with *claim* and correlates with *cov_hgih*. When we further observe their conditional correlation, we use the dummy variables (*highspeed_1* and *highspeed_2*) which represents the individual whose *avspd* and *maxspd* are higher than medium. We find that *maxspd* is not anymore a good choice for significant factor in representing individual's risk or risk preference because the dummy variable *highspeed_2* is not anymore a significant factor when we observe from its conditional correlation with *claim*. However, *avspd* is still a good choice, the conditional correlation of *highspeed_1* is significant when it is correlated with *claim* and when it is correlated with *cov_high*. It is worth to be mentioned that, *highspeed_1* represents for individual's risk preference because the sign of its conditional correlation with *claim* and with *cov_high* are opposite.

Although the correlation between *ratio_night* and *cov_high* is insignificantly different from 0 no matter observe from simple correlation or observe from conditional correlation, we still keep it as one the control variable in our model for the purpose of comparing with Kremslehner and Muermann (2016).

Till now, in additional to the characteristic variables of insured and insured vehicle which are collected from the underwriting information in the insurance companies, the variables which represent for the driving information are also chosen. We list the descriptive statistics of all these variables related to the driving information in panel B of Table 3, and the descriptive statistics of all these variables related to the characteristics of insured and insured vehicle in Table 6.

[Insert Table 6 Here]

4 Methodology

The aim of this research is to explore the private information recorded in the telematics of the insured vehicles. We adopt the bivariate Probit regressions model which is also adopted in Chiappori and Salanie (2000).

The bivariate Probit regressions model is as following:

$$claim_i = 1(X_i\beta_X + telem_i\beta_{telem} + \epsilon_i^{clm} > 0) \tag{1}$$

$$cov_high_i = 1(X_i\beta_X + telem_i\beta_{telem} + \epsilon_i^{cov} > 0) \tag{2}$$

In regression 1 and regression 2, vector X_i contains all variables which are used by the insurance company to underwrite and price this insurance contract. $telem_i$ is the vector which contains the chosen variables which indicate the information recorded from telematics.

In bivariate Probit regressions model, there is a correlation coefficient (ρ) between the residuals (ϵ_i^{clm} and ϵ_i^{cov}) of the above two regressions. According to the traditional concept, we judge that whether there is asymmetric information by whether the null hypothesis of $\rho = 0$ is rejected and infer the asymmetric information as adverse selection/moral hazard when $\rho$is positive, whereas infer that the asymmetric information as advantageous selection when$\rho$is negative.

After Finkelstein and McGarry (2006), the concept of multiple dimensions of asymmetric information is raised. The new approach to test asymmetric information is that, even when the correlation coefficient of the two residuals of the bivariate probit regressions (ρ) is significantly different from 0 and positive (negative), or insignificantly different from 0, the advantageous selection (adverse selection/moral hazard) still existed when there is variable which represents the private information of the insured, and both of the estimated coefficients of this variable in two regressions are with the opposite (same) signs to each other. In this research, we focus the private information recorded from the telematics. If the private information represents the individual's risk in driving, we expect that both estimated coefficients of this variable in the two regressions are significantly different from 0 and with the same sign. If the private information represents the individual's risk preference which is emerged in driving behavior, we expect that both the estimated coefficients of this variable in two regressions are significantly different from 0 and with the opposite sign.

As mentioned in previous section, the information collected from the telematics is classified into three categories. When we focus on the private information recorded from the telematics, we choose some representative variables from each category. *N_rides* and *avmlg* are chosen as the representative variables in driving amount. Radical driving behavior which is indicated by three dummy variables: *radical_1*, *radical_2*, and *radical_3*, is the representative in driving habit. High speed driver (*highspeed_1*) is chosen as the representative of individual's risk preference.

In addition to the above variables are chosen by us from the information recorded in telematics, we also hold the information that the driver's ratio of over speed (*speeding*) and the ratio of driving at night (*ratio_night*). Both of these two variables are also adopted by Kremslehner and Muermann (2016). In order to compare our empirical outcomes with theirs, these two variables are also included in the explanatory variables in our research.

Hence, by adopting bivariate probit regressions, we build our Model 1 when *speeding*, *N_rides*, *ratio_night* , *avmlg* are included in vector *telem*. Our Model 1 is close to the model of Kremslehner and Muermann (2016) as benchmark. According to the preliminary observation in last section, *speeding N_rides*, *ratio_night* , and *avmlg* are indices of driver's traffic risk. Hence, we predict that the estimated coefficients of them should be positive in both regressions.

In our research sample, *radical_1*, *radical_2*, and *radical_3* are also similar indices to *speeding* to indicate the driver's radical driving behavior. Hence, in Model 2 we replace the *speeding* in Model 1 by *radical_1*, *radical_2*, and *radical_3*. Except for the predictions in Model 1, we also predict that the estimated coefficients of *radical_1*, *radical_2*, and *radical_3* should be positive in both regressions.

In Model 3, we include all categories of driving information which we have chosen in the vector *telem*. The chosen variables of driving information are: *highspeed_1*, *radical_1* , *radical_2*, *radical_3* *N_rides*, *ratio_night*, and *avmlg*. Except for the prediction on *radical_1*, *radical_2*, *radical_3* *N_rides*, *ratio_night*, and *avmlg*, we further predict that the estimated coefficients of *highspeed_1* should be with opposite sign when *highspeed_1* is in the regression which regresses on *claim*, and when *highspeed_1* is in the regression which regresses on *cov_high* because we predict that *highspeed_1* is an index for individual's risk preference.

In each model, the key variable to be observed is the estimated coefficient β_{telem} in regression (1) and in regression (2). If both the two estimated coefficients β_{telem} in regression (1) and (2) are significantly different from 0, we can confirm that the information included in vector *telem* is useful private information. Even when the ρ of two residuals in the two regressions is insignificantly different from 0, the asymmetric information problem could still exist. As for what kind of asymmetric information it is or they are, we have to judge by the sign of the estimated coefficient of the private information variable when its estimated coefficient is significantly different from 0. Multiple dimensions of asymmetric information problem is especially worth to be noticed. When valuable private information variables show both same direction of correlation and different direction of correlation with *claim*/*cov_high* in the meantime, or when the valuable private information variable shows a different way of asymmetric information problem with that shown by the correlation coefficient of the residuals of bivariate probit model.

5 Empirical Results

As previously discussed, Kremslehner and Muermann (2016) investigate the insured who have purchased UBI contracts, whereas we investigate the insured who are observed but not yet covered by the UBI contract. Because whether there is the concern of price penalty from UBI contract or not could potentially affect the behaviors of driving, the first task in our research is to compare the empirical results between ours and that of Kremslehner and Muermann (2016) by a similar model.

[Insert Table 7 Here]

Model 1 is the similar model to provide this comparison as reported in Table 7. We find that the estimated coefficients of variable *N_rides* in regression (1) and in regression (2) are both significantly different from 0 and positive. Although the estimated coefficients of variable *avmlg* are both insignificantly different from 0, *avmlg* is only used for controlling the driving distance, our results show the significant explanatory ability of driving amount. The drivers who drive more tend to have higher traffic accident risk. This finding can also be observed from Kremslehner and Muermann (2016). When the distance driven has been controlled for, they also find that the drivers who drive more tend to have higher traffic risk. And both of our and their researches find that these high risk drivers also tend to purchase higher coverage of insurance contracts.

However, when the estimated coefficients of *speeding* are both significantly different from 0 and positive in regression (1) and in regression (2), the estimated coefficients of their "*AvgSpeeding*" is insignificant factor of risk in Kremslehner and Muermann (2016). This implies that the over speed driving behavior is observed and estimated as a significant risk driving behavior by our research, but this phenomenon is not sustained in Kremslehner and Muermann (2016). This distinct difference between our research and theirs confirms our doubt that the risky driving behavior of the insured under UBI could be restrained than only be observed. UBI contract have the actual effect on lower the effect of risky driving behavior of the insured. As for the ratio of driving at night (*ratio_night*), we find that it has no significant effect on traffic risk. This is consistent with the finding in

Kremslehner and Muermann (2016).

Furthermore, we observe that the correlation coefficient in this bivariate probit model is negative and significantly different from 0 at 1% of significance level. When the private information *speeding*, and *N_rides* have shown significantly positive correlation with *claim* and *cov_high* in the meantime, we cannot infer that only advantageous selection exited by merely judging from the correlation coefficient as tradition. The unobservable factors are related with *claim* and *cov_high* oppositely. Meanwhile, some private information revealed by driving behavior indicate individual's risk type are related with *claim* and *cov_high* in the same direction. There are multiple dimensions of asymmetric information problem existed. In contrast, when some variables related to the driving information are significantly correlated with *claim* and *cov_high* in the meantime in Kremslehner and Muermann (2016), they cannot reveal any information to help to infer multiple dimensions of asymmetric information. It is because that those variables related to the driving information are not "private information" to the insurance company. They have been used for underwriting and pricing in UBI contract.

[Insert Table 8 Here]

In Model 2 as reported in Table 8, we further try to use the index which can indicate the "radical driving behavior" to represent risky driving behavior. Although driving fast is risky, dramatically rapid acceleration and/or dramatically rapid deceleration is even radical driving behavior than just driving fast. When we replace *speeding* by *radical_1*, *radical_2*, *radical_3*, we find that, comparing to the one who would neither frequently rapid deceleration nor frequently rapid acceleration, the one who frequently rapid acceleration and rapid deceleration and the one who frequently rapid acceleration although not frequently rapid deceleration are more risky drivers. They are much more likely to have traffic accidents, the estimated coefficients of *radical_1*, *radical_2* are positive and significantly different from 0 in regression (1). They also tend to purchase higher coverage contract because they know they are high risk, the estimated coefficients of *radical_1*, *radical_2* are also positive and significantly different from 0 in regression (2). As for the rest empirical outcomes of other variables are basically consistent with the outcomes in Model 1.

Worth to be noticed, *radical_1* and *radical_2* are much more significant and influential than *speeding*. So, the contribution of our Model 2 is that we find: the radical driving behaviors observed by whether they frequently rapid acceleration and rapid deceleration is much powerful than observe the speeding behavior when we want use it as a factor to indicate the driving risk. Furthermore, just frequently rapid deceleration is also risky factor, but it is not as significant as "rapid acceleration" as well as "rapid deceleration". However, when we observe that the estimated coefficients of *radical_3* are not significantly different from 0 in regression (1) as well as in regression (2), which means that if the drivers only frequently rapid deceleration, but not often rapid acceleration, they are not so risky. Does this observation imply that rapid acceleration is one kind of initiative behavior, and rapid deceleration is only passive behavior? Can the driving risk only be pushed by initiative radical driving behavior instead of being affected by some passive reactions such as rapid deceleration? This is worthy to be further explored in the future.

The multiple dimension of asymmetric information problems are also proved to be emerged in Model 2. *N_rides*, *radical_1*, and *radical_2* are significantly positive correlated with *claim* and *cov_hig*h in the same time. Meanwhile, the correlation coefficient of residuals in bivariate probit model is negative and significantly different from 0.

[Insert Table 9 Here]

Model 3 is the one which contains most complete information of driving as reported in Table 9. Three categories of driving information are included in this most complete model. *N_rides* and *avmlg* are the variables represent the driving amount. *radical_1*, *radical_*2, *radical_3* are the variables represent the driving habit. *highspeed_1* represents the risk preference of the individual. The empirical results of Model 3, which are in Table 9, are basically consistent with Model 2 for those variables also existed in Model 2. The only additional variable in Model 3 is *highspeed_1*. We can see that its estimated coefficient in regression (1) is positive and its estimated coefficient in regression (2) is negative. Both of them are significantly different from 0. So, it is treated as the individual's risk preference index: the individuals who tend to drive fast (higher than the medium of average driving speed) are less risk aversion. Hence, they tend to have higher probability of accident and tend to less willing to purchase high coverage insurance contract. Worth to be mentioned, this *highspeed_1* shows to be in an opposite direction of explanation when it isn't put together with other driving information in the regressions. Hence, this phenomenon reminds us that the information controlled in the model should be as complete as possible. Otherwise, the spurious correlation could arise because of the missing variable problem.

In Model 3, we can once again observe the private information related to risk type (such as: *N_rides*, *radical_1*, and *radical_2*) and the private information related to risk preference (such as: *highspeed_1* and the significantly negative correlation coefficient of residuals) coexisted when the correlation coefficient of residuals in bivariate probit regressions model is negative and significantly different from 0. Multiple dimensions of asymmetric information problems are the inference again.

6 Conclusion

Since the asymmetric information is a persistent important problem for the insurance markets, exploring meaningful private information and using it in underwriting and pricing is very important. Through the new technology of telematics, the private information of driving behaviors can be revealed by the big data of telematics, and the UBI contract in automobile insurance can be developed accordingly.

In this paper, we find that several driving behaviors are valuable private information which reveals either individual's risk or individual's risk preference. The driving amount indicated by the number of rides per day is a significant factor of driving risk. The one who drives more tends to have higher driving risk and also tends to purchase higher coverage of insurance. Radical driving behaviors such as speeding or frequently rapid acceleration/deceleration are also indices of driving risk. The one who often commit speeding or has more frequent rapid acceleration/deceleration than others tends to have higher driving risk and purchase high coverage contract. As an important contribution, our proposed new index of frequently rapid acceleration/deceleration is more powerful indicator than speeding behavior.

We also identify several driving behaviors as indices of risk preference. For example, we find that the one who drives faster than others is less risk aversion, and tends to have higher driving risk and less willing to purchase high coverage contract. In addition, we provide the indirect evidence that risky driving behavior can be restrained by UBI contract. Our research help answer the question that whether an individual will have different reaction when they are merely observed or when they have the concern of price penalty in insurance. We also confirm the prior literature that the market exists with multiple dimension of asymmetric information instead of single dimension of asymmetric information.

In summary, we are among the very first research to use the driving behavior information recorded by the telematics to analyze the asymmetric information problem in automobile insurance market. We show that several driving behaviors are valuable private information which reveals either individual's risk or individual's risk preference. In addition, we find evidence of adverse selection/moral hazard associated with the private information revealed from big data. Our proposed research provide new insights to help the insurance company better understand the big data from telematics and therefore better align driving behaviors with premium rates for auto insurance.

References

Ackerloff, G., 1970. The market for lemons: Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 84(3), pp.488-500.

Ayuso, M., Guillen, M. and Perez-Marin, A. M., 2016, Telematics and Gender Discrimination: Some usage-based evidence on whether men's risk of accident differs from women's, Risks, 4 (10): 1-10.

Bond, E.W. and Crocker, K.J., 1991. Smoking, skydiving, and knitting: the endogenous categorization of risks in insurance markets with asymmetric information. Journal of Political Economy, pp.177-200.

Bouvard, M., 2014. Real option financing under asymmetric information. Review of Financial Studies, 27(1), pp.180-210.

Brockett, P.L. and Golden, L.L., 2007. Biological and psychobehavioral correlates of credit scores and automobile insurance losses: Toward an explication of why credit scoring works. Journal of Risk and Insurance, 74(1), pp.23-63.

Cardon, J.H. and Hendel, I., 2001. Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey. RAND Journal of Economics, pp.408-427.

Cohen, A., & Siegelman, P. (2010). Testing for adverse selection in insurance markets. Journal of Risk and Insurance, 77(1), 39-84.

Cutler, D. M., & Reber, S. J. (1998). Paying for health insurance: the trade-off between competition and adverse selection. The Quarterly Journal of Economics, 113(2), 433-466.

Donnelly, C., Englund, M., Nielsen, J.P. and Tanggaard, C., 2014. Asymmetric information, self-selection, and pricing of insurance contracts: The simple no-claims case. Journal of Risk and Insurance, 81(4), pp.757-780.

Einav, L. and Finkelstein, A., 2010, A graphical analysis of selection in insurance markets, Working paper.

Finkelstein, A. and McGarry, K., 2006, Multiple dimensions of private information: Evidence from the long-term care insurance market, American Economic Review 96(4): 938–958.

Finkelstein, A. and Poterba, J., 2014. Testing for asymmetric information using "unused observables" in insurance markets: Evidence from the UK annuity market. Journal of Risk and Insurance, 81(4), pp.709-734.

Kremslehner, D. and Muermann A., 2016, Asymmetric information in automobile insurance: evidence from driving behavior, working paper.

Harris, M., & Raviv, A. (1978). Some results on incentive contracts with applications to education and employment, health insurance, and law enforcement. The American Economic Review, 68(1), 20-30.

Hendren, N., 2013. Private information and insurance rejections. Econometrica, 81(5), pp.1713-1762. Litman, T., 2005, Pay-as-you-drive pricing and insurance regulatory objectives, Journal of Insurance Regulation, 23 (3): 35-53.

Litman, T., 2012, Disrance-based vehicle insurance as a TDM strategy, Victoria Transport Policy Institute.

Spindler, M., Winter, J. and Hagmayer, S., 2014. Asymmetric information in the market for automobile insurance: Evidence from Germany. Journal of Risk and Insurance, 81(4), pp.781-801.

Roberts, M.R., 2015. The role of dynamic renegotiation and asymmetric information in financial contracting. Journal of Financial Economics, 116(1), pp.61-81.

Rothschild, M. and Stiglitz, J., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The quarterly journal of economics, pp.629-649.

Wang, K. C., Peng, J. L., Sun, Y. Y., and Chang, Y. C., 2011, The Asymmetric Information Problem in Taiwan's Cancer Insurance Market, The Geneva Risk and Insurance Review. 36, 202-219.

Zavadil, T., 2015. Do the better insured cause more damage? Testing for asymmetric information in car insurance. Journal of Risk and Insurance, 82(4), pp.865-889.

Table 1: Summary of variables and their definitions

Variable	Definition
cov_high	A dummy variable which equals 1 when the amount of the third party liability insurance the insured purchased is over 0.3 million RMB, otherwise it equals 0.
claim	A dummy variable which equals 1 when there is claim filed during the policy year, otherwise it equals 0.
ins_PD	A dummy variable which equals 1 when the insured further purchased the first party's property damage insurance contract, otherwise it equals 0.
bm	The number represents the bonus-malus coefficient according to the claim records in last three policy years.
seat_under6	A dummy variable which equals 1 when the vehicle's seats is less than 6, otherwise it equals 0.
teleph	A dummy variable which equals 1 when the contract of the insured is from telephone marketing, otherwise it equals 0.
emply_gov	A dummy variable which equals 1 when the insured is government's employee, and 0 otherwise.
emply _corp	A dummy variable which equals 1 when the insured is company's employee, otherwise it equals 0.
company_A	A dummy variable which equals 1 when the insurance contract is sold by company A, otherwise it equals 0.
company_B	A dummy variable which equals 1 when the insurance contract is sold by company B, otherwise it equals 0.
female	A dummy variable which equals 1 when the insured is female, otherwise it equals 0.
age_under20	A dummy variable which equals 1 when the insured is less 20 years old, otherwise it equals 0.
age_2025	A dummy variable which equals 1 when the insured is or older than 20 and is younger than 25, otherwise it equals 0.
age_2530	A dummy variable which equals 1 when the insured is or older than 25 and is younger than 30, otherwise it equals 0.
age_3060	A dummy variable which equals 1 when the insured is or older than 30 and is younger than 60, otherwise it equals 0.
carage_under3	A dummy variable which equals 1 when the age of the inured vehicle is less than 3 years, otherwise it equals 0.
carage_3-6	A dummy variable which equals 1 when the age of the inured vehicle is or is over 3 years and less than 6 years, otherwise it equals 0.
carage_6-9	A dummy variable which equals 1 when the age of the inured vehicle is or is over 6 years and less than 9 years, otherwise it equals 0.
loyalty	A dummy variable which equals 1 when the insured continue to purchase the contract from the same insurance company at least for one year, otherwise it equals 0.
N_rides	It represents the average number of rides per driving day.
avmlg	It represents the average driving mileage per driving day
time_drv	It represents the average time of driving per driving day.
rate_use	It represents the percentage of driving days during our research period.
freq_acc	It represents the frequencies of rapid acceleration per driving day.
freq_dec	It represents the frequencies of rapid deceleration per driving day.
radical_i (i=1,2,3)	A dummy variable that equals 1 when the insured is type i radical driver, otherwise it equals 0.
avspd	It represents the average driving speed per driving day.
maxspd	It represents the level of maximum driving speed.
highspeed_1	A dummy variable that equals 1 when the average driving speed (avspd) of the insured is kept above the medium, otherwise it equals 0.
highspeed_2	A dummy variable that equals 1 when the maximum driving speed (maxspd) of the insured is kept above the medium, otherwise it equals 0.
speeding	It represents the ratio of over speed, which is measured by the time of speeding over the time of driving.
ratio_night	It represents the ratio of driving at night, which is measured by the time of driving at night over the time of driving.

Note: Type 1 radical driver is defined as the insured whose frequencies of rapid acceleration per day (freq_acc) are higher than the level of 25% quantile, and whose frequencies of rapid deceleration (freq_dec) are higher than the level of 50% quantiles. Type 2 radical driver is defined as the insured whose frequencies of rapid acceleration per day (freq_acc) are higher than the level of 25% quantile, but whose frequencies of rapid deceleration (freq_dec) are lower than the level of 50% quantiles. Type 3 radical driver is defined as the insured whose frequencies of rapid acceleration per day (freq_acc) are lower than the level of 25% quantile, but whose frequencies of rapid deceleration (freq_dec) are higher than the level of 50% quantiles.

Table 2: The structure of insurance amount

Coverage (RMB)	Percentage	Cumulative Percentage
50,000	1.09	1.09
100,000	7.3	8.39
150,000	0.24	8.63
200,000	17.05	25.68
300,000	25.23	50.91
500,000	39.29	90.2
1,000,000	9.67	99.87
1,500,000	0.06	99.94
2,000,000	0.06	100

Table 3: The summary statistics of driving information

Variable	Q_1	Q_2	Q_3
Panel A: raw information			
N_rides	3.57	4.41	5.7
avmlg	23.476	32.975	47.194
time_drv	1.53	1.97	2.63
rate_use	0.6154	0.7941	0.907
avspd	13.714	15.9506	18.722
speeding	0	0.0002	0.0025
maxspd	72.7	82.021	93.206
freq_acc	0.11	0.46	1.49
freq_dec	0.44	0.85	1.64
ratio_night	0.0594	0.1048	0.1661
Panel B: the variables related to driving used in our research			
Variable	Mean		Standard Deviation
N_rides	4.9562		2.1349
avmlg	39.3098		24.1771
radical_1	0.4347		0.4957
radical_2	0.3075		0.4615
radical_3	0.0636		0.244
highspeed_1	0.4999		0.5
speeding	0.0033		0.0078
ratio_night	0.1226		0.0862

Table 4: Correlation matrix

	claim	*cov_hgih*	*N_rides*	*avmlg*	*time_drv*	*speeding*	*freq_acc*	*freq_dec*	*avspd*	*maxspd*	*ratio_night*
claim	1	-0.0219	0.0727	0.0598	0.0746	0.0305	0.0337	0.0456	-0.0204	0.0349	0.0295
		0.0537	<.0001	<.0001	<.0001	0.0072	0.003	<.0001	0.0725	0.0021	0.0092
cov_hgih	-0.0219	1	0.0812	0.0633	0.0385	0.057	0.0518	0.0296	0.0697	0.1211	-0.0058
	0.0537		<.0001	<.0001	0.0007	<.0001	<.0001	0.0089	<.0001	<.0001	0.6102
N_rides	0.0727	0.0812	1	0.6471	0.7028	0.0917	0.1406	0.3157	-0.0468	0.274	0.2047
	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
avmlg	0.0598	0.0633	0.6471	1	0.8963	0.2547	0.1426	0.368	0.3434	0.3964	0.28
	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
time_drv	0.0746	0.0385	0.7028	0.8963	1	0.0569	0.1243	0.3665	-0.0583	0.2317	0.2919
	<.0001	0.0007	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
speeding	0.0305	0.057	0.0917	0.2547	0.0569	1	0.1166	0.0931	0.4726	0.514	0.032
	0.0072	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	0.0048
freq_acc	0.0337	0.0518	0.1406	0.1426	0.1243	0.1166	1	0.4906	0.0597	0.3126	0.1072
	0.003	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
freq_dec	0.0456	0.0296	0.3157	0.368	0.3665	0.0931	0.4906	1	0.0662	0.2271	0.1318
	<.0001	0.0089	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
avspd	-0.0204	0.0697	-0.0468	0.3434	-0.0583	0.4726	0.0597	0.0662	1	0.4562	-0.0173
	0.0725	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	0.1277
maxspd	0.0349	0.1211	0.274	0.3964	0.2317	0.514	0.3126	0.2271	0.4562	1	0.1296
	0.0021	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001
ratio_night	0.0295	-0.0058	0.2047	0.28	0.2919	0.032	0.1072	0.1318	-0.0173	0.1296	1
	0.0092	0.6102	<.0001	<.0001	<.0001	0.0048	<.0001	<.0001	0.1277	<.0001	

Note: In each line of each variable, the value in upper row is the correlation coefficient (ρ) between two corresponding variables; the value in lower row is the p value for the test of H0:$\rho = 0$.

Table 5: The estimated coefficients for each single driving behavior variable in bivariate probit regressions

Parameter Estimates				
	Claim		Coverage	
Parameter	Estimate	P value	Estimate	P value
N_rides	0.0483	<.0001	0.0432	<.0001
avmlg	0.0035	<.0001	0.0027	<.0001
time_drv	0.0686	<.0001	0.0241	0.013
speeding	5.2463	0.0196	8.0767	<.0001
freq_acc	0.0112	0.0296	0.0216	<.0001
freq_dec	0.0273	0.0005	0.0205	0.0026
highspeed_1	-0.0627	0.0994	0.1486	<.0001
highspeed_2	0.0309	0.4217	0.2214	<.0001
ratio_night	0.4592	0.0319	-0.1487	0.374

Table 6: Summary statistics

Variable	Mean	Std Dev
Claim	0.1136	0.3173
cov_high	0.4909	0.5000
ins_PD	0.9562	0.2047
Bm	0.8846	0.1430
seat_under6	0.9326	0.2508
Teleph	0.1905	0.3927
emply_gov	0.3872	0.4871
emply _corp	0.3058	0.4608
company_A	0.5368	0.4987
company_B	0.3058	0.4608
female	0.2723	0.4452
age_under20	0.0004	0.0196
age_2025	0.0202	0.1406
age_2530	0.1562	0.3631
age_3060	0.7905	0.4070
carage_under3	0.7103	0.4536
carage_3-6	0.223	0.4163
carage_6-9	0.0551	0.2282
loyalty	0.4343	0.4957

Table 7: The empirical results of Model 1

	Claim		Coverage	
Parameter	Estimate	P value	Estimate	P value
Intercept	-1.7892	<.0001	-0.4022	0.033
ins_PD	-0.775	<.0001	-0.455	<.0001
Bm	0.5564	0.0003	-0.1485	0.1929
seat_under6	-0.0696	0.3827	0.0782	0.1766
Teleph	0.0996	0.1224	-0.1698	0.0003
emply_gov	0.0049	0.9452	-0.1082	0.0409
emply _corp	0.2512	0.4831	0.0411	0.8915
company_A	-0.0692	0.3648	0.0904	0.1136
company_B	0.0308	0.5951	0.0232	0.6032
female	0.0517	0.2235	-0.0354	0.2758
age_under20	0.5847	0.4575	5.0341	0.9728
age_2025	0.174	0.2755	0.3217	0.0132
age_2530	-0.0221	0.8477	0.3722	<.0001
age_3060	-0.047	0.6594	0.2688	0.0011
carage_under3	-0.1956	0.3202	0.234	0.0966
carage_3-6	-0.1485	0.4535	-0.0236	0.8678
carage_6-9	-0.1836	0.3851	-0.1101	0.4662
loyalty	-0.0351	0.4317	0.0496	0.1388
speeding	4.11	0.079	7.2107	0.0002
N_rides	0.0412	0.0002	0.0433	<.0001
ratio_night	0.1958	0.3857	-0.3718	0.0332
avmlg	0.0007	0.5196	0.0001	0.9004
ρ		-0.0653 (0.0062)		

Note: ρ is the correlation coefficient between the two residuals in bivariate probit regressions. The number in the parentheses is the p value to test the null hypothesis that the correlation coefficient of the two residuals is positive.

Table 8: The empirical results of Model 2

	Claim		Coverage	
Parameter	Estimate	P value	Estimate	P value
Intercept	-1.8717	<.0001	-0.5079	0.0076
ins_PD	-0.7813	<.0001	-0.4597	<.0001
bm	0.5631	0.0003	-0.1433	0.2093
seat_under6	-0.0598	0.4556	0.0951	0.1011
teleph	0.0964	0.1352	-0.1735	0.0002
emply_gov	0.0027	0.9702	-0.1116	0.0351
emply _corp	0.2266	0.5267	0.0027	0.9929
company_A	-0.0672	0.3792	0.0946	0.0978
company_B	0.0333	0.5651	0.0261	0.5592
female	0.0518	0.2232	-0.0364	0.2616
age_under20	0.588	0.4514	5.033	0.9742
age_2025	0.1756	0.2705	0.3296	0.011
age_2530	-0.0157	0.8916	0.3867	<.0001
age_3060	-0.0373	0.7267	0.2871	0.0005
carage_under3	-0.1916	0.3293	0.2481	0.0785
carage_3-6	-0.1586	0.4225	-0.0262	0.8535
carage_6-9	-0.1949	0.3558	-0.1147	0.4484
Loyalty	-0.0331	0.4584	0.053	0.1141
radical_1	0.1849	0.0012	0.2064	<.0001
radical_2	0.1091	0.0614	0.1433	0.0006
radical_3	0.0515	0.5714	0.0229	0.7315
N_rides	0.0356	0.0013	0.0351	0.0001
ratio_night	0.1525	0.4999	-0.4361	0.0125
avmlg	0.0005	0.5984	0.0004	0.6516
ρ		-0.0682 (0.0042)		

Note: ρ is the correlation coefficient between the two residuals in bivariate probit regressions. The number in the parentheses is the p value to test the null hypothesis that the correlation coefficient of the two residuals is positive.

Table 9: The empirical results of Model 3

	Claim		Coverage	
Parameter	Estimate	P value	Estimate	P value
Intercept	-1.9243	<.0001	-0.4084	0.0331
ins_PD	-0.7823	<.0001	-0.4586	<.0001
bm	0.5642	0.0003	-0.1478	0.1954
seat_under6	-0.05	0.5334	0.0785	0.1772
teleph	0.0958	0.1377	-0.1711	0.0003
emply_gov	0.0011	0.988	-0.1093	0.0392
emply _corp	0.2155	0.5457	0.0363	0.9041
company_A	-0.0658	0.3892	0.092	0.1077
company_B	0.0329	0.57	0.0259	0.5633
Female	0.0496	0.2435	-0.0313	0.3356
age_under20	0.6417	0.411	4.9312	0.9745
age_2025	0.1963	0.2192	0.2931	0.0243
age_2530	0.0015	0.9895	0.3567	<.0001
age_3060	-0.0293	0.7838	0.274	0.0009
carage_under3	-0.1843	0.3478	0.2333	0.0986
carage_3-6	-0.155	0.4331	-0.0348	0.8063
carage_6-9	-0.1876	0.3739	-0.1292	0.3943
Loyalty	-0.0331	0.4581	0.0533	0.1124
highspeed_1	0.0803	0.048	-0.1586	<.0001
radical_1	0.1942	0.0007	0.1905	<.0001
radical_2	0.1116	0.0559	0.1397	0.0009
radical_3	0.0508	0.5766	0.0254	0.7028
N_rides	0.0295	0.0103	0.0482	<.0001
ratio_night	0.1194	0.5987	-0.3712	0.034
avmlg	0.0012	0.2776	-0.001	0.2392
ρ		-0.0658 (0.0060)		

Note: ρ is the correlation coefficient between the two residuals in bivariate probit regressions. The number in the parentheses is the p value to test the null hypothesis that the correlation coefficient of the two residuals is positive.