Can Vehicle Maintenance Records Predict Automobile Accidents?

Shyi-Tarn Bair
CEO, Ho-An Insurance Agency CO., LTD, Taiwan

Rachel J. Huang*
Associate Professor, Graduate Institute of Finance
National Taiwan University of Science and Technology, Taipei, Taiwan

Kili C. Wang
Associate Professor, Department of Insurance
Tamkang University, Taipei, Taiwan

*Corresponding Author. E-mail address: rachel@mail.ntust.edu.tw

Can Vehicle Maintenance Records Predict Automobile Accidents?

Abstract

This paper proposes that vehicle maintenance is one of the factors affecting the probability of automobile accidents. To test the hypothesis, we use a unique data set which is merged from an insurance company and a vehicle manufacturer in Taiwan. The total number of observations in our final sample is 155,116. We find that proper maintenance defined by the recommended kilometers is significantly negatively correlated with loss probability in compulsory automobile liability insurance. On average, maintenance will decrease the loss probability by between 0.165% and 0.261% according to different proxies for proper maintenance, whereas the average loss probability for the overall sample is 0.93%. We further find that proper maintenance is insignificantly correlated with loss severity.

JEL classification: G22, C12, D01
Keywords: Maintenance Record; Automobile Insurance; Accident

1. Introduction

Understanding the factors that affect automobile accidents is an essential and important issue due to the fact that automobile insurance is the most important insurance line in many countries. For example, in the United States, personal automobile insurance is the largest line of insurance in the Property and Casualty insurance industry. The total net written premiums for personal automobile insurance amounted to 159.7 billion dollars in the year 2007, accounting for a share of 35.7% of the Property and Casualty insurance. The average growth rate of the net written premiums for personal automobile insurance was about 4.2% per year from the year 2000 to 2007.[1]

A large body of literature has focused on the determinant factors predicting automobile accidents. In addition to the main classificatory variables used in the automobile insurance market, such as gender, age, marital status, zip code and the claim history, researchers have further found that credit history (Tillman and Hobbs, 1949; Miller and Smith, 2003), the market price of an automobile (Andersson, 2005), and drinking and sensation seeking behavior (such as driving fast and racing cars) (Iversen and Rundmo, 2002) are correlated with automobile accidents.

Recently, Brockett and Golden (2007) provide a detailed survey of the related literature and classify the risk factors to biological, psychological and behavioral characteristics. More importantly, they hypothesize that the reason that credit scoring works in underwriting automobile insurance is that there is an underlying biological and/or psychological and social component of each individual which regulates their risk taking behavior, and, since this propensity is individual and intrinsic, it spans risk

[1] The data are obtained from the U.S. Census Bureau, Statistical Abstract of the United States: 2010, Table 1185.

scenarios, from financial decision making to risky driving decision making.

Inspired by Brockett and Golden (2007), this paper intends to study whether maintenance records can be another variable to present the biological and/or psychological and social component of each individual, and could therefore be correlated to automobile accidents. Specifically, in this paper, we empirically investigate whether maintenance records can predict the automobile accident rates and automobile accident claim amounts.

Why can maintenance records influence automobile accidents? First, following Brockett and Golden (2007)'s hypothesis, people who are more responsible in their vehicle maintenance might be also more responsible in their driving behavior. Second, the vehicle maintenance records could be a signal of the degree of risk aversion of the driver. Jindapon and Neilson (2007) show that a more risk-averse individual is willing to expend more efforts on self-protection for a mean-preserving improvement in the risk. If the car is properly maintained, then the driver may be a more risk-averse individual. Therefore, on the one hand, she may have less of a chance of being involved an automobile accident. On the other hand, she may not cause a severe automobile accident even if she is involved in one. Third, if the car is properly maintained, then, for mechanical reasons, the frequency and severity of automobile accidents may be lower.

Based on the above arguments, we set up two hypotheses. The first one is that proper vehicle maintenance can reduce the probability of an automobile accident. The second hypothesis is that proper vehicle maintenance can decrease the loss severity. To test these two hypotheses, we use a unique data set which is merged from an insurance company and a vehicle manufacturer. Our data is representative since both companies have the largest market shares in their respective industries in Taiwan. The written

premium of the insurance company was about 20% of the automobile insurance market in 2008, whereas the market share of the car dealer was about 35% in 2008.

The data from the insurance company allows us to observe claim records, insurance contracts and all variables used by the insurance company in underwriting and pricing from the year 2001 to the year 2006. The data from the vehicle manufacturer contain the maintenance dates, kilometers traveled, items involved in the vehicle maintenance from the year 2000 to 2006. Only the observations which are recorded by both the manufacturer's repair shop and the insurance company are included in our final sample. In total, we have 155,116 observations.

The insurance we focus on is compulsory automobile liability insurance, which is designed to protect the third parties' life and body from the damage caused by the usage of motor vehicles. There are three reasons why we explore the determining factors in predicting automobile accidents within this market. First, third party liability is the coverage most under the control of the insured. Therefore, the biological and psychological characteristics of the insured might have a major impact on this line of insurance. Second, we will have the largest sample size possible because each vehicle must by law participate in this insurance. Third, the claim records will serve as a good proxy for accidents in this market. According to Chiappori and Salanie (2000), using claim records as a proxy for accidents may suffer from a bias of un-claimed accidents. However, this problem is less serious because there is no deductible and there must have been a third party involved.

Two econometric models are used in this paper to examine our hypotheses. We first use the Probit regression to test the relationship between the loss probability and vehicle maintenance. Second, by employing samples where the individuals have at some point

filed a claim, we use OLS regression to investigate whether maintenance records could explain the claim amounts of third party liability insurance.[2]

We find that the maintenance records are negatively correlated with the occurrence of compulsory automobile liability claims. To be specific, the individuals who maintain their cars according to the recommended number of kilometers have a significantly lower accident rate. The average claim rate for the whole sample is 0.93% per year. The average loss probability decreases by 0.165% when the insured vehicle is properly maintained according to the recommended number of kilometers in the previous year. If the vehicle is properly maintained according to the recommended number of kilometers in the current year, then proper maintenance will reduce the loss probability by 0.202%. When proper maintenance is defined by where the maintenance is done according to the recommended number of kilometers each time it is maintained during the current and the previous years, proper maintenance will decrease the loss probability by 0.261% in average. The empirical evidence supports the view that not only can the maintenance records predict the occurrence of the risk but they are also essential factors that influence the probability of a claim.

As for the relationship between proper maintenance and loss severity, we find that the maintenance records are not significant factors in terms of predicting the claim amounts for compulsory automobile liability. Although proper maintenance will decrease the loss severity through the use of different proxies for proper maintenance, the relationship is statistically insignificant.

The remainder of this paper is organized as follows. In Section 2, we provide a brief introduction to the automobile insurance market in Taiwan. Section 3 describes our data.

[2] To keep as many observations as possible, we do not adopt panel data techniques to analyze our sample in the first place. We will further construct a panel data to examine our hypothesis in the section of robustness analyses.

Section 4 introduces our main empirical methodologies. Section 5 reports our empirical findings. Section 6 provides robustness analyses. In Section 6, we construct a balanced panel data and adopt panel data technology to examine our hypotheses. We also provide sensitivity analysis by varying the threshold while defining proper maintenance. We further examine the relationship between maintenance records and the insured's characteristics. Section 7 concludes the paper.

2. The Automobile Liability Insurance Market in Taiwan

As in many other countries in the world, there are two systems of automobile liability insurance in Taiwan: compulsory and voluntary. Each motor vehicle must be insured under the system of compulsory liability, which is designed to protect the third party's life as well as protect the third party against bodily injury caused by the usage of the vehicle. Individuals can also purchase voluntary third party liability insurance, which covers bodily injury or property damage sustained by the third party, to compensate in the event of insufficient coverage under compulsory insurance.

There are three types of payment in regard to compulsory insurance: life, maiming and medical expenses. The compulsory insurance coverage for life is NT$1,500,000.[3] The degree of incapacity is divided into 15 levels. If the injured third party becomes disabled as the result of an accident, he or she will be reimbursed by between NT$40,000 and NT$1,500,000 according to the level of incapacity. The medical expenses include the costs of first aid and treatment. The upper-limit indemnity for the medical expenses of the injured third party is NT$200,000.

In 2008, a total of 13,177,754 vehicles were covered by compulsory insurance.

[3] The average exchange rate in 2008 was approximately NT$31.54 to US$1, and the standard deviation was NT$1.09 per US dollar.

Written premiums in 2008 amounted to NT$17,704 million and exhibited an average growth rate of 1.08% since the year 2001. The average loss ratio was 71.6% between 2001 and 2008.[4]

3. Data

Our empirical data consist of two main components. One consists of the data of compulsory automobile liability insurance, while the other is concerned with the vehicles' maintenance records. It is worth noting that we only include one particular brand instead of different brands of cars in our final sample.[5] This is because motor vehicle manufacturers only take care of their own brand of vehicles. Although the use of a brand name imposes certain limitations, the total number of observations in our final sample of 155,116 is still fairly large during the sample period from 2001 to 2006. The size of our sample is not small due to the fact that both companies are the largest ones in their respective industries.

The definitions of our variables are listed in Table 1. The two dependent variables used in the two different empirical models include a dummy variable of whether the insured files a claim (noted by *claim*) and the logarithm of the claim amount[6]

[4] In Taiwan, the average total net written premiums for personal automobile insurance between 2000 and 2006 was 483 billion NT dollars each year, which accounted for around 50% of the Property and Casualty insurance market on average. The average annual growth rate of the net written premium of personal automobile insurance was about 1.79% over this period. The above data were obtained from the website of the Taiwan Insurance Institute (http://www.tii.org.tw/).

[5] Although our sample only contains one brand of cars, our sample has similar feature to the observations that purchase compulsory liability from our insurance company. Specifically, our sample has insignificant differences with the data from the sample insurance company with respect to individual's age, marriage status and the vehicle registration location. However, our data has significant higher percentage of female and new cars (age one and two) observations. The individuals in our data also have higher probability of purchasing insurance policies through the channel of a car dealer-owned agency. We further compare our sample with Li, Liu and Yeh's (2007) data. They use the entire market data of comprehensive policies in Taiwan from 2000 to 2001. Our data show insignificant differences in the percentage of gender and new cars with theirs.

[6] The claim record in relation to the compulsory automobile liability insurance is the proxy for the occurrence of the risk.

($\ln amount$). The explanatory variables in this research include the characteristics of the insured individuals and insured cars which are listed in the part for the independent variables in Table 1.

The key control variable that we focus on is the one that indicates whether the vehicle is properly maintained. We adopt different proxies for it. In practice, the vehicle owners can visit the repair shops of the vehicle manufacturer to maintain their car according to the recommended number of kilometers or dates.[7] The sample vehicle manufacturer suggests that vehicles should visit be taken to their service and repair centers every 10,000 kilometers or every 6 months. If the recommended number of kilometers or date is missing, the repair shop can still follow the maintenance instructions to replace or check the suggested items when the vehicle is sent back to the shop. Thus, we use three criteria to identify whether the vehicle has been properly maintained.

The first criterion, vm_km , evaluates proper maintenance according to whether the maintenance is carried out according to the recommended number of kilometers during the year. Note that we allow an extra 20% in regard to the recommended number of kilometers.[8] Therefore, if the vehicle is sent in for maintenance every 12,000 kilometers in year t , then $vm\_km_t = 1$, otherwise $vm\_km_t = 0$. The second criterion, vm_time , evaluates proper maintenance according to whether the maintenance is conducted according to the recommended times during the year. Again, an additional 20% of the recommended time is allowed. Thus, if in year t the vehicle is sent for maintenance within 7.2 months of the previous maintenance record, then $vm\_time_t = 1$, otherwise

[7] The vehicle maintenance records cover only those repairs and maintenance that is performed at the brand dealership, so it may understate the results and make the observed significance less than it actually is (i.e., the statistical significance found is conservative). This choice of information source for the vehicle maintenance records, while inevitable practically, may mean, for example, that older used cars and younger drivers are underrepresented.

[8] We will further relax the extra 20% threshold in the Section of robustness analyses.

$vm\_time_t = 0$. The third criterion, vm_item, evaluates proper maintenance according to whether the maintenance is conducted according to the recommended items each time during the year. If the maintenance record in year t shows that the vehicle is checked according to the maintenance instructions or more than required by those instructions, then $vm\_item_t = 1$, otherwise $vm\_item_t = 0$. The above three criteria for proper maintenance are calculated according to the maintenance record for year t, $t-1$ and both years (denoted by a suffix "$2years$").

The other control variables are listed in Table 1. They include the underwriting variables of the compulsory automobile liability insurance, such as the individual's age and gender. We also include the other variables to control the heterogeneity for the insured, such as the individual's marriage status, the vehicle's age, size, model, the registration area, the sales channel of the insurance policy, and the year dummies.

The basic statistics for the final sample are listed in Table 2. The probability of having filed a compulsory liability insurance claim is around 0.93%. Most of the insured who have filed a claim file only one such claim during a policy year. About 20 of them file two claims, only one of them files three claims, and none of them file more than three claims during a policy year.

Based on the index for vm_km_t, 11.99% of insured cars are properly maintained; according to the index for vm_time_t, 15.63% of insured cars are properly maintained; and based on the index for vm_item_t, 14.23% of insured cars are properly maintained.[9] Both vm_time_t and vm_item_t are slightly higher than vm_km_t. This may arise because both vm_time_t and vm_item_t indicate that some of the car owners maintain

[9] One might wonder that cars involved in accidents may not be maintained on time simply because of their accident involvement. Our data shows that, among the claimed sample, only 1.44%, 0.90%, and 1.26% of the observations face this kind of problem when the proper maintenance is defined according to kilometers, time, and item, respectively.

their cars only because they are recommended to do so,[10] while vm_km_t indicates that the car owners know when their cars need to be maintained by themselves observing the number of kilometers of usage. The car maintenance status for the previous year also reveals the same information. Both the indexes for vm_km_{t-1} and vm_item_{t-1} are, respectively, close to vm_km_t and vm_item_t, and are around 10%. The index for vm_time_{t-1} is higher than that for vm_time_t and reaches 22%.

Most of the insured are married and between 30 and 60 years old, the percentage of insured in this age range is over 80%, and the percentage of married insured is 92.3%. Female insured account for over half of the total, and the percentage is 69%. One of the reasons why most vehicles are owned by married middle-aged females is due to the insurance premium discounts for them. The insurance premium is dependent on the characteristics of the owner of the vehicle rather than the drivers in Taiwan. Since married middle-aged females can enjoy a premium discount, most families register their vehicle under the name of a married middle-aged female member of the family.

The percentage of cars under three years old in Taiwan is around 60%. About 50% of insured cars are used in cities, in the northern area of Taiwan, and most (over 90%) of them are non-commercial sedans. The bigger cars (over 2000 cc.) account for about 26% of the total. Around 60% of the insurance contracts in our empirical sample were sold by the car dealers' own agents.

In our empirical work, we further investigate whether the cars' maintenance records are able to explain the severity of third-party liability risk. Therefore, we constrain our data into a sub-sample of those insured who have filed a claim. The basic statistics for the sub-sample of the claimed insured are listed in Table 2. The average claim amount is

[10] The vehicle manufacturer from which we obtained our data will often remind its customers to maintain their vehicles when the recommended maintenance time is up. They also advise which items will be maintained each time the vehicle is maintained.

NT$92,374.[11] The structures of some characteristics, e.g., sex, age, marital status of the insured, and the age, usage of the insured cars, remain the same as in the full sample. All of the maintenance indices are lower in this sub-sample than in the full sample. This may imply that proper maintenance could predict the risk of third-party liability, and the ones who properly maintain their vehicle may have a lower risk probability.

4. Methodology

There are two major hypotheses in this paper that may be described as follows:

Hypothesis 1: The accident probability will decrease when the insured vehicle is properly maintained.

Hypothesis 2: The accident severity will decrease when the insured vehicle is properly maintained.

To test the first hypothesis, a Probit regression is employed. The Probit regression is as follows:

$$\text{Prob}(claim_i = 1 \mid vm_i, X_i) = \Phi(vm_i\beta + X_i\gamma) \quad (1)$$

where $claim_i = 1$ when the insured has filed the claim on compulsory automobile liability insurance during the policy year, otherwise $claim_i = 0$. Φ is the cumulative distribution function of $N(0,1)$, β and γ are the coefficients for the explanatory variables vm_i and X_i. The vector X_i contains all the underwriting variables and year dummy variables which are listed in Table 1. The variable vm_i denotes the maintenance records during the policy year for vehicle i. There are three kinds of indexes to record maintenance. Hence, we construct four models. In Model 1, we only use vm_km as the maintenance record variable. Only vm_time is used in Model 2 and only vm_item is

[11] The maximum value of the claim amount is NT$1,839,967. The minimum value of the claim amount is NT$4,060.

used in Model 3. As for Model 4, we use all of these three indexes.

The prediction ability of the maintenance record in regard to the loss probability is captured by the significance of the estimated coefficient $\hat{\beta}$. The predicted sign of $\hat{\beta}$ is negative since our hypothesis is that the risk of third-party liability is negatively correlated with the maintenance records. We further calculate the change in estimated claim probability due to proper maintenance.

To test our second hypothesis, we set up another regression model which involves regressing on the claim amount of the compulsory automobile liability insurance. The regression model is:

$$\ln amount_i = vm_i\alpha + X_i\delta + \varepsilon_i, \qquad (2)$$

where $\ln amount_i$ denotes the logarithm of the claim amount. Again, we use three different proxies for vm_i to create four models as they are in Equation (1). We also vary the time periods for the vehicle maintenance records for each model. The key coefficient that we focus on is the significance of the estimated coefficient $\hat{\alpha}$. We also predict that the sign of $\hat{\alpha}$ should be negative if the hypothesis of negative correlation between risk severity and maintenance is sustained.

5. Empirical Results

The empirical results regarding the relationship between loss probability and maintenance records are presented from Tables 3 to 5. We find that all proxies for proper maintenance are negatively correlated with risk probability, which supports our first hypothesis.

As shown in Table 3, when the previous maintenance records are used to estimate the loss probability in the current year, only the coefficient of vm_km_{t-1} is significantly

negative among those three indexes of maintenance. However, if these three indexes are simultaneously used as in Model 4, the coefficient of vm_km_{t-1} becomes insignificant. When the current maintenance records are used to test the current loss probability, Table 4 shows that both the coefficients of vm_km_t and vm_item_t are significantly negatively correlated with the loss probability. If these three indexes are simultaneously used as in Model 4, the coefficient of vm_km_t is still significantly negative. Furthermore, when both current and previous maintenance records are used, the results (as shown in Table 5) are consistent with those in Table 4. These findings show that regardless of whether the vehicle is proper maintained according to the number of kilometers of usage, i.e., vm_km, is valuable when it comes to estimating the loss probability.

It is worth noting that vm_time is insignificant in all models, although it is also negatively correlated with loss probability. One of the reasons may be that this variable reveals less information related to the individual's private characteristics, such as risk aversion. Our sample vehicle manufacturer will call its customers to remind them of the scheduled maintenance when the recommended date is approaching. Thus, it is not clear whether the driving force behind a vehicle maintained according to the suggested time is the individual's characteristics or the reminder.

Consistent with the findings in the literature, Tables 3 to 5 demonstrate that young people (those under 25 years old) and females tend to have a higher loss probability,[12] and that new cars have a higher loss probability than old cars (over 4 years old). In addition, we find that the loss probability is lower in the city than in the suburban areas. This might be because traffic flows speed more slowly in the city areas.

We further calculate the average reduction in loss probability from maintenance in

[12] Note that the individuals are the owners of the vehicle. The owner and the driver of the vehicle may not be the same one.

Models 1 and 4. As shown in Table 6, the average loss probability decreases by 0.165% when the insured vehicle is properly maintained according to the recommended number of kilometers in the previous year. If the vehicle is properly maintained according to the recommended number of kilometers in the current year, then proper maintenance will reduce the loss probability by 0.202%. When proper maintenance is defined by where the maintenance is done according to the recommended number of kilometers each time it is maintained during the current and the previous years, proper maintenance will decrease the loss probability by 0.261% in average. Similar results are obtained when Model 4 is used. Compared to the average loss probability, 0.93%, the reduction in loss probability due to vehicle maintenance is substantial.

The results of regressing the logarithm of the claim amount as shown in regression equation (2) are listed from Tables 7 to 9. The results demonstrate that the estimated coefficients of the *vm* variable in each regression are insignificant. These imply that the information regarding whether the car owners properly maintain their car or not can not provide valuable information on predicting the severity of risk for those who have already filed a claim.

It should be noted that, in addition to *vm* , there are several variables that have significant explanatory power in terms of loss probability but not loss severity. Only the variables *year*2004 and *north* have significant explanatory power in terms of both loss probability and severity. The data for the year 2004 has a higher loss probability and loss severity than that for year 2001. The insured vehicles registered in north Taiwan have significantly lower loss probability and loss severity than the vehicles registered in middle Taiwan.

Tables 7 to 9 further show that the old cars (over 4 years) have a higher loss severity

than younger cars. However, the reduction in loss severity is not significant for the cars within two years. We also find that married vehicle owners have a significantly higher loss severity than single owners. This finding does not consist with standard insurance heuristics. One possible explanation[13] could be that, due to particularities of the Taiwan insurance market, the women are the registered owners when there are married people, even when the men are overwhelmingly driving (due to cultural effects), while the single drivers have their own sex, which may be roughly equal in frequency. Thus, if men have a higher accident severity, then they contribute to most of the exposure of married driving but only half the exposure of the single drivers.

6. Robustness Analyses

In this section, we provide several robustness analyses. First, we construct a sub-sample with balanced panel data to examine the effect of proper maintenance on loss probability and loss severity. Second, we focus on sensitive analyses. We vary the threshold while defining proper maintenance. Third, we examine the factors that affect the individual's vehicle maintenance behavior.

6.1. Panel Analysis

In our data set, some of the drivers are included for several years. Therefore, we could construct a panel data to further examine the relationship between proper maintenance and accidents. To conduct a balanced panel data, we track the individuals who are in our sample during 2004 and 2006 rather than during 2000 and 2006 and obtain 22,977 observations. Comparing the basic statistics in the balanced sub-sample (Table 10)

[13] We thank the anonymous referee for providing the possible explanation.

with that in the full sample (Table 2), we find a lower average claim probability and a higher percentage of proper maintenance in the balanced sub-sample.

Table 11 reports the coefficients and the p-value of the proper maintenance variables.[14] Panel A in Table 11 shows that proper maintenance defined by the recommended number of kilometers or recommended items could reduce liability claim probability. However, the effect is statistically significant only when two-year maintenance records are used to define proper maintenance. Consistent with the findings in the previous section, Panel B in Table 11 shows that proper maintenance does not have significant impact on loss severity.

6.2. Sensitivity Analyses

We vary the chosen threshold (20% extra time/miles) while defining proper maintenance to examine the sensitivity of the relationship between proper maintenance and loss probability.[15] We allow an extra 10% and 15% in regard to the recommended number of kilometers and recommended time.

Table 12 reports the coefficients and the p-value of the proper maintenance variables of Models 1, 2 and 4 in the Probit regression of full sample. It shows that no matter which threshold is chosen, proper maintenance according to the recommended number of kilometers could reduce the probability of having a compulsory liability claim. Interestingly, the drivers might face a significant higher loss probability when the vehicle is properly maintained according to the recommended time. The evidence is significant when an extra 10% and 15% threshold is chosen by using the maintenance records in

[14] All of the estimations in this section were carried out using the random effects model.

[15] The sensitivity of the relationship between proper maintenance and loss severity is also examined. The relationship is consistently insignificant by using different threshold. Since the results are insignificant and consistent to the previous findings, we do not report them. The results could be obtained from the authors upon request.

current year and both current and previous years. Note that the variable vm_time might not be a good proxy for the driver's degree of risk aversion or "responsibility", since it is also affected by the manufacturer's behavior.

We further use the balanced panel sub-sample to do sensitivity analyses. Table 13 reports the results of random effect model by varying the threshold of extra recommended kilometers from 10%, 15% to 20% when defining the variable $vm\_km_{2\,years}$. The results in Table 13 again confirm our previous findings that a driver who maintains his vehicle according to the recommended kilometers will face a lower loss probability. Note that, regarding to other control variables, we find that only $city$ is significant by using panel data.

6.3. The characteristics of the individual whose vehicle is properly maintained

The above findings imply that the vehicle maintenance records might be a signal of the insured's characteristics. Therefore, we further adopt a Probit regression to examine the relationship between maintenance records and the insured's characteristics. Specifically, the factors which impact the decision to maintain the insured vehicle properly according to the recommended number of kilometers (vm_km) are examined. Thus, we adopt the following Probit regression:

$$\text{Prob}(vm\_km_i = 1 \mid X_i) = \Phi(X_i\theta), \tag{3}$$

where θ is the coefficient for the explanatory variable X_i, which contains all the independent variables listed in Table 1.

The first three columns of Table 14 demonstrate the impact of the insured's characteristics on the previous year, current year and two years maintenance records. The results show that senior drivers (age 60 and up) have higher probability to properly

maintain their vehicle than junior drivers (below age 25). In average, senior people is more risk averse than young people. This finding provides partial evidence that more risk averse individuals would have a higher probability to invest in risk prevention.

In addition, one-year-old new cars have higher probability to be properly maintained. Usually, new cars have a higher value than old cars. This finding is consistent to the argument that people have a higher probability to protect their property when the value of the property is higher. Further, we find that if the driver lives in the city area, then the probability of proper maintenance is lower. This might reflect the competition of vehicle service centers in city areas. In Taiwan, there are many small and convenient vehicle service centers in city areas which provide efficient and simple maintenance services. Some of the individuals live in city might choose to maintain their vehicle at these convenient vehicle service shops instead of at the manufacturer's service centers.

Since the maintenance behavior is determined by the individual's characteristics, one might wonder that proper maintenance could be endogenous. In other words, we might need to control the endogenous problem when examining the relationship between claim probability and proper maintenance. To take care of this problem, we further calculate the estimated probability of proper maintenance in Equation (3), which is denoted by $Prob(vm\_km)$, and add it in Equation (1) as a control variable. The results are shown in the last three columns of Table 14. Comparing these results with those of Model 1 in Tables 3, 4 and 5, we find that the proper maintenance variable itself is still significantly negatively correlated with the claim probability. The coefficients of other variables also show similar patterns after controlling the estimated probability of proper maintenance. These evidences suggest that our findings in the previous section are robust.

7. Conclusion

Much of the literature has explored the determining factors in terms of predicting automobile accidents. Recently, the focus of risk classification factors for automobile accidents has been on the biological, psychological and behavioral characteristics of human beings. Inspired by the literature, this paper intends to explore new valuable risk prediction factors based on the individual's bio-psychological behavior. We examine whether or not maintenance records can predict automobile accidents.

We find that maintenance records do contain valuable information in predicting automobile accidents. We evaluate proper maintenance based on three different criteria: kilometers, time, and items maintained. We find that vehicle proper maintenance measured by kilometers is significantly negatively correlated with loss probability by using the current year or both current and previous years maintenance records. The loss probability of a vehicle properly maintained according to the recommended number of kilometers is reduced by between 0.165% and 0.261% according to the different periods for maintenance records. Although we find that vehicle maintenance can reduce loss probability, we cannot reject the hypothesis that maintenance records are not correlated with loss severity.

Our paper contributes to the literature by linking the vehicle maintenance records with automobile liability claims. We provide further support to Brockett and Golden's (2007) hypothesis in a way that people who are more responsible in their vehicle maintenance are also more responsible in their driving behavior, even after adjusting for standard underwriting variables. Brockett and Golden (2007) supply the biological and psychological underpinnings from the finances and responsibility with credit domain, whereas we provide additional evidence in vehicle maintenance domain.

There are some extensions that could be performed in future studies. First, our data are limited to a particular vehicle brand. Therefore, collecting more data for different brands of vehicle might be useful to better understand the relationship between vehicle maintenance, loss probability and loss severity. Second, we only focus on liability automobile insurance. The relationship between vehicle maintenance and automobile property insurance is also worth examining. Third and moreover, both credit scoring and maintenance records tap the underlying dimension of the biological and/or psychological and social component of individuals. If collecting both types of information is possible, it would have certain contribution to the credit scoring and insurance debate by examining the relationship between credit scores and vehicle maintenance.

References

Andersson, H., 2005, The Value of Safety as Revealed in the Swedish Car Market: An Application of the Hedonic Pricing Approach, ***The Journal of Risk and Uncertainty***, 30:3; 211–239.

Brockett, P. L., and L. L. Golden, 2007, Biological and Psychobehavioral Correlates of Credit Scores and Automobile Insurance Losses: Toward An Explication of Why Credit Scoring Works, ***The Journal of Risk and Insurance***, 74:1; 23-63.

Chiappori, P.A., and B. Salanie, 2000, Testing for Asymmetric Information in Insurance Markets, ***Journal of Political Economy,*** 108: 1; 56-78.

Iversen, H., and T. Rundmo, 2002, Personality, Risky Driving and Accident Involvement Among Norwegian Drivers, ***Personality and Individual Differences***, 33; 1251-1263.

Jindapon, P., and W. S. Neilson, 2007, Higher-order Generalizations of Arrow–Pratt and Ross Risk Aversion: A Comparative Statics Approach, ***Journal of Economic Theory***, 136; 719 – 728

Miller, M., and R. A. Smith, 2003, The Relationship of Credit-based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity, ***Actuarial Study***, Epic Actuaries online at http://www.epicactuaries.com

Li, C., C. Liu and J. Yeh, 2007, The Incentive Effects of Increasing Per-Claim Deductible Contracts in Automobile Insurance, ***Journal of Risk and Insurance***, 74; 441-459.

Tillman, W. A., and G. E. Hobbs, 1949, The Accident-Prone Automobile Driver: A Study of the Psychiatric and Social Background, ***The American Journal of Psychiatry***, 106; 321-331.

Table 1 Definitions of the variables

Variables	Definition
Dependent variables:	
$claim$	A dummy variable that equals 1 when the insured has filed an at fault claim on the compulsory automobile third party liability insurance, otherwise it equals 0.
$\ln amount$	Logarithm of the claim amount of the insured for the whole policy year.
Independent variables:	
Vehicle maintenance records:	
vm_km_{t-1}	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended number of kilometers each time it is maintained during year *t-1*, otherwise it equals 0.
vm_time_{t-1}	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended time each time it is maintained during year *t-1*, otherwise it equals 0, otherwise it equals 0.
vm_item_{t-1}	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended items each time it is maintained during year *t-1*, otherwise it equals 0.
vm_km_{t}	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended number of kilometers each time it is maintained during year *t*, otherwise it equals 0.
vm_time_{t}	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended time each time it is maintained during year *t*, otherwise it equals 0.
vm_item_{t}	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended items each time it is maintained during year *t*, otherwise it equals 0.
$vm\_km_{2\,years}$	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended number of kilometers each time it is maintained during both the previous and the current year, otherwise it equals 0.
$vm\_time_{2\,years}$	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended time each time it is maintained during both the previous and the current year, otherwise it equals 0.
$vm\_item_{2\,years}$	A dummy variable that equals 1 when the vehicle maintenance is done according to the recommended items each time it is maintained during both the previous and the current year, otherwise it equals 0.
Control variables:	
$age2530$	A dummy variable that equals 1 when the insured is between the ages of 25 and 30, otherwise it equals 0.
$age3060$	A dummy variable that equals 1 when the insured is between the ages of 30 and 60, otherwise it equals 0.
$age60up$	A dummy variable that equals 1 when the insured is over the age of 60, otherwise it equals 0.[a]
$female$	A dummy variable that equals 1 when the owner of the car is female, otherwise it equals 0.
$married$	A dummy variable that equals 1 when the owner of car is married, otherwise it equals 0.
$carage1$	A dummy variable that equals 1 when the car is one year old, otherwise it equals 0.
$carage2$	A dummy variable that equals 1 when the car is two years old, otherwise it equals 0.
$carage3$	A dummy variable that equals 1 when the car is three years old, otherwise it equals 0.
$carage4$	A dummy variable that equals 1 when the car is four years old, otherwise it equals 0.[b]
big	A dummy variable that equals 1 when the insured car equals or is over 2000 c.c., otherwise it equals 0.
$sedan$	A dummy variable that equals 1 when the car is a sedan and is for non-commercial or for long-term rental purposes, otherwise it equals 0.[c]
$channel\_D$	A dummy variable that equals 1 when the policy is sold through the channel of a car dealer-owned agency, otherwise it equals 0.
$city$	A dummy variable that equals 1 when the owner of the car lives in a city, otherwise it equals 0.
$north$	A dummy variable that equals 1 when the car is registered in the north of Taiwan, otherwise it equals 0.
$south$	A dummy variable that equals 1 when the car is registered in the south of Taiwan, otherwise it equals 0.

east	A dummy variable that equals 1 when the car is registered in the east of Taiwan, otherwise it equals 0[d].
*year*2002	A dummy variable that equals 1 when the data belong to the year 2002, otherwise it equals 0.
*year*2003	A dummy variable that equals 1 when the data belong to the year 2003, otherwise it equals 0.
*year*2004	A dummy variable that equals 1 when the data belong to the year 2004, otherwise it equals 0.
*year*2005	A dummy variable that equals 1 when the data belong to the year 2005, otherwise it equals 0.
*year*2006	A dummy variable that equals 1 when the data belong to the year 2006, otherwise it equals 0.

Note: a. The reference group for the dummy variables related to age includes the insured who are under 25 years old.

b. The reference group for the dummy variables related to the car age is that which includes all the cars used over four years.

c. The reference group for the dummy variable includes the insured cars which are for non-commercial use, other than sedans for long-term rental purposes.

d. The reference group for those three dummy variables related to area includes the cars registered in central Taiwan.

Table 2 Basic statistics of full sample

	All observations		Claimed observations	
Variables	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
claim	0.0093	0.0958		
amount [e]			92,374	223,537
vm_km_{t-1}	0.1155	0.3196	0.0974	0.2966
vm_time_{t-1}	0.2232	0.4164	0.2088	0.4066
vm_item_{t-1}	0.1363	0.3431	0.1190	0.3239
vm_km_{t}	0.1199	0.3248	0.0974	0.2966
vm_time_{t}	0.1563	0.3631	0.1420	0.3491
vm_item_{t}	0.1423	0.3494	0.1232	0.3288
$vm\_km_{2\,years}$	0.0547	0.2275	0.0404	0.1969
$vm\_time_{2\,years}$	0.1279	0.3340	0.1148	0.3189
$vm\_item_{2\,years}$	0.0723	0.2590	0.0598	0.2373
*age*2530	0.0628	0.2425	0.0564	0.2307
*age*3060	0.8888	0.3144	0.8949	0.3068
*age*60*up*	0.0407	0.1977	0.0362	0.1868
female	0.6924	0.4615	0.7481	0.4343
married	0.9231	0.2664	0.9193	0.2725
*carage*1	0.3788	0.4851	0.4078	0.4916
*carage*2	0.2685	0.4432	0.2693	0.4438
*carage*3	0.1654	0.3715	0.1601	0.3668
*carage*4	0.1014	0.3019	0.0939	0.2919
big	0.2609	0.4391	0.2199	0.4143
sedan	0.9826	0.1309	0.9861	0.1172
channel_D	0.6062	0.4886	0.6354	0.4815
city	0.5202	0.4996	0.4426	0.4969
north	0.5160	0.4997	0.3786	0.4852
south	0.2722	0.4451	0.3410	0.4742
east	0.0206	0.1421	0.0306	0.1723
*year*2002	0.0902	0.2865	0.0877	0.2829
*year*2003	0.1327	0.3393	0.1364	0.3433
*year*2004	0.1742	0.3793	0.1976	0.3984
*year*2005	0.2402	0.4272	0.2589	0.4382
*year*2006	0.3121	0.4634	0.2728	0.4455
Number of observations	155116		1437	

Note: e. The variable *amount* is the total claim amount of the insured who has filed the claim for the whole policy year. The variable does not use logarithms here.

Table 3 Probit regression of compulsory liability claim (using the previous year's maintenance record)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	-2.2808	<.0001	-2.2887	<.0001	-2.2815	<.0001	-2.2795	<.0001
vm_km_{t-1}	-0.0705	0.0386					-0.0678	0.1696
vm_time_{t-1}			-0.0104	0.7007			-0.0023	0.9328
vm_item_{t-1}					-0.0491	0.1193	-0.0029	0.9488
*age*2530	-0.2216	0.0306	-0.2219	0.0303	-0.2228	0.0296	-0.2218	0.0305
*age*3060	-0.1592	0.0974	-0.1600	0.0957	-0.1607	0.0941	-0.1594	0.0970
*age*60*up*	-0.1776	0.1018	-0.1799	0.0973	-0.1796	0.0979	-0.1778	0.1014
female	0.0714	0.0017	0.0710	0.0018	0.0713	0.0017	0.0713	0.0017
married	-0.0283	0.4564	-0.0286	0.4515	-0.0287	0.4510	-0.0283	0.4564
*carage*1	0.1041	0.0133	0.1015	0.0166	0.1016	0.0155	0.1045	0.0137
*carage*2	0.0694	0.1051	0.0691	0.1067	0.0683	0.1102	0.0695	0.1049
*carage*3	0.0516	0.2519	0.0515	0.2529	0.0512	0.2559	0.0516	0.2518
*carage*4	0.0351	0.4745	0.0350	0.4752	0.0349	0.4763	0.0351	0.4746
big	-0.0658	0.0056	-0.0655	0.0058	-0.0656	0.0057	-0.0658	0.0056
sedan	0.0769	0.3480	0.0749	0.3598	0.0756	0.3558	0.0767	0.3488
channel_D	0.0188	0.3891	0.0282	0.2226	0.0213	0.3306	0.0178	0.4580
city	-0.0799	<.0001	-0.0794	<.0001	-0.0798	<.0001	-0.0799	<.0001
north	-0.2033	<.0001	-0.2030	<.0001	-0.2031	<.0001	-0.2032	<.0001
south	-0.0284	0.2903	-0.0285	0.2874	-0.0293	0.2750	-0.0284	0.2895
east	0.0209	0.7409	0.0194	0.7582	0.0207	0.7425	0.0210	0.7398
*year*2002	0.0349	0.5414	0.0350	0.5407	0.0360	0.5286	0.0349	0.5411
*year*2003	0.0682	0.2051	0.0684	0.2041	0.0706	0.1897	0.0683	0.2047
*year*2004	0.1138	0.0295	0.1136	0.0299	0.1164	0.0261	0.1139	0.0295
*year*2005	0.1031	0.0436	0.1039	0.0420	0.1059	0.0382	0.1032	0.0437
*year*2006	0.0287	0.5737	0.0304	0.5512	0.0314	0.5381	0.0288	0.5731
χ^2 *value of Likelihood Ratio*	187.8893	<0.0001	183.6391	<0.0001	185.9595	<0.0001	187.9006	<0.0001

Table 4 Probit regression of compulsory liability claim (using the current year's maintenance record)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	-2.2757	<.0001	-2.2901	<.0001	-2.2761	<.0001	-2.2680	<.0001
vm_km_t	-0.0826	0.0148					-0.1036	0.0930
vm_time_t			-0.0102	0.7326			-0.0263	0.3942
vm_item_t					-0.0579	0.0627	0.0161	0.7785
*age*2530	-0.2219	0.0302	-0.2216	0.0305	-0.2232	0.0293	-0.2219	0.0303
*age*3060	-0.1592	0.0973	-0.1596	0.0964	-0.1609	0.0937	-0.1594	0.0970
*age*60*up*	-0.1778	0.1014	-0.1797	0.0977	-0.1796	0.0978	-0.1783	0.1004
female	0.0716	0.0016	0.0709	0.0018	0.0715	0.0017	0.0711	0.0018
married	-0.0289	0.4478	-0.0286	0.4519	-0.0290	0.4463	-0.0286	0.4519
*carage*1	0.1021	0.0150	0.1000	0.0173	0.1009	0.0162	0.1046	0.0129
*carage*2	0.0693	0.1052	0.0687	0.1085	0.0693	0.1052	0.0702	0.1012
*carage*3	0.0520	0.2487	0.0514	0.2536	0.0518	0.2500	0.0521	0.2472
*carage*4	0.0352	0.4730	0.0351	0.4743	0.0348	0.4780	0.0355	0.4694
big	-0.0663	0.0053	-0.0654	0.0059	-0.0659	0.0055	-0.0657	0.0057
sedan	0.0759	0.3537	0.0750	0.3591	0.0752	0.3586	0.0754	0.3568
channel_D	0.0165	0.4469	0.0295	0.1822	0.0189	0.3882	0.0095	0.6915
city	-0.0797	<.0001	-0.0793	<.0001	-0.0799	<.0001	-0.0796	<.0001
north	-0.2031	<.0001	-0.2032	<.0001	-0.2032	<.0001	-0.2025	<.0001
south	-0.0275	0.3054	-0.0287	0.2852	-0.0277	0.3010	-0.0279	0.2979
east	0.0214	0.7347	0.0194	0.7590	0.0209	0.7405	0.0221	0.7265
*year*2002	0.0346	0.5454	0.0351	0.5388	0.0352	0.5376	0.0345	0.5459
*year*2003	0.0685	0.2035	0.0684	0.2035	0.0689	0.2006	0.0684	0.2042
*year*2004	0.1128	0.0311	0.1138	0.0295	0.1139	0.0294	0.1125	0.0315
*year*2005	0.1016	0.0469	0.1043	0.0412	0.1029	0.0441	0.1015	0.0471
*year*2006	0.0285	0.5762	0.0307	0.5474	0.0293	0.5653	0.0284	0.5775
χ^2 *value of Likelihood Ratio*	189.6314	<0.0001	183.6081	<0.0001	187.0265	<0.0001	190.5297	<0.0001

Table 5 Probit regression of compulsory liability claim (using both the previous and current years' maintenance records)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	-2.2843	<.0001	-2.2894	<.0001	-2.2830	<.0001	-2.2779	<.0001
$vm\_km_{2\,years}$	-0.1244	0.0114					-0.1250	0.0548
$vm\_time_{2\,years}$			-0.0153	0.6343			-0.0237	0.4654
$vm\_item_{2\,years}$					-0.0722	0.0826	-0.0035	0.9500
*age*2530	-0.2206	0.0313	-0.2216	0.0305	-0.2226	0.0297	-0.2210	0.0311
*age*3060	-0.1575	0.1011	-0.1598	0.0961	-0.1603	0.0948	-0.1582	0.0998
*age*60*up*	-0.1754	0.1062	-0.1798	0.0974	-0.1789	0.0992	-0.1763	0.1045
female	0.0712	0.0017	0.0708	0.0018	0.0713	0.0017	0.0708	0.0018
married	-0.0285	0.4537	-0.0286	0.4522	-0.0290	0.4464	-0.0284	0.4563
*carage*1	0.1027	0.0145	0.1009	0.0165	0.0997	0.0175	0.1054	0.0124
*carage*2	0.0689	0.1073	0.0689	0.1075	0.0681	0.1115	0.0697	0.1034
*carage*3	0.0520	0.2481	0.0515	0.2531	0.0515	0.2533	0.0522	0.2466
*carage*4	0.0355	0.4691	0.0350	0.4745	0.0350	0.4756	0.0356	0.4682
big	-0.0665	0.0051	-0.0653	0.0059	-0.0660	0.0055	-0.0659	0.0055
sedan	0.0761	0.3524	0.0747	0.3611	0.0749	0.3604	0.0752	0.3584
channel_D	0.0216	0.3075	0.0289	0.1858	0.0241	0.2581	0.0163	0.4708
city	-0.0801	<.0001	-0.0793	<.0001	-0.0800	<.0001	-0.0800	<.0001
north	-0.2036	<.0001	-0.2030	<.0001	-0.2033	<.0001	-0.2031	<.0001
south	-0.0276	0.3035	-0.0287	0.2852	-0.0284	0.2894	-0.0279	0.2988
east	0.0207	0.7428	0.0193	0.7597	0.0208	0.7418	0.0211	0.7387
*year*2002	0.0350	0.5406	0.0351	0.5388	0.0361	0.5280	0.0351	0.5392
*year*2003	0.0691	0.1997	0.0685	0.2030	0.0705	0.1905	0.0693	0.1985
*year*2004	0.1138	0.0295	0.1138	0.0294	0.1161	0.0265	0.1140	0.0294
*year*2005	0.1030	0.0440	0.1043	0.0412	0.1052	0.0395	0.1031	0.0437
*year*2006	0.0290	0.5694	0.0307	0.5470	0.0311	0.5417	0.0291	0.5682
χ^2 *value of Likelihood Ratio*	190.3183	<0.0001	183.7186	<0.0001	186.6078	<0.0001	190.8554	<0.0001

Table 6 The decreased claim probability due to proper maintenance

Variables	Claim probability with proper maintenance (1)	Claim probability without proper maintenance (2)	Decreased claim probability (2)-(1)
Panel A: Model 1			
vm_km_{t-1}	0.0078006***	0.0094551***	0.0016545***
	(0.002868)	(0.003384)	(0.000515)
vm_km_{t}	0.0074903***	0.0095056***	0.0020153***
	(0.002789)	(0.003387)	(0.000598)
$vm\_km_{2\,years}$	0.0067978***	0.0094068***	0.0026090***
	(0.002516)	(0.003880)	(0.000865)
Panel B: Model 4			
vm_km_{t-1}	0.0078005***	0.0094551***	0.0016546***
	(0.003846)	(0.003846)	(0.000542)
vm_km_{t}	0.0074881***	0.0095060***	0.0020179***
	(0.002674)	(0.003415)	(0.000741)
$vm\_km_{2\,years}$	0.0067963***	0.0094069***	0.0026106***
	(0.002516)	(0.003385)	(0.000869)

Note: *** denotes statistical significance at the 1-percent. The value in the parenthesis is their standard deviations.

Table 7 Regression of compulsory liability claim amount in the claimed sub-sample (using the previous year's maintenance record)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	10.5270	<.0001	10.5114	<.0001	10.5292	<.0001	10.5122	<.0001
vm_km_{t-1}	-0.0207	0.8981					-0.0143	0.9491
vm_time_{t-1}			0.0323	0.7981			0.0362	0.7776
vm_item_{t-1}					-0.0219	0.8832	-0.0184	0.9288
*age*2530	-0.4528	0.3151	-0.4518	0.3161	-0.4546	0.3130	-0.4507	0.3181
*age*3060	-0.3951	0.3466	-0.3928	0.3496	-0.3970	0.3444	-0.3925	0.3507
*age*60*up*	-0.3783	0.4316	-0.3796	0.4297	-0.3802	0.4291	-0.3777	0.4329
female	-0.1532	0.1536	-0.1536	0.1527	-0.1534	0.1531	-0.1536	0.1530
married	0.3726	0.0347	0.3725	0.0347	0.3723	0.0348	0.3731	0.0346
*carage*1	-0.2007	0.3156	-0.2108	0.2950	-0.2004	0.3162	-0.2076	0.3041
*carage*2	-0.1850	0.3651	-0.1893	0.3543	-0.1846	0.3665	-0.1872	0.3605
*carage*3	-0.4454	0.0378	-0.4482	0.0364	-0.4452	0.0379	-0.4462	0.0377
*carage*4	-0.5476	0.0182	-0.5485	0.0180	-0.5471	0.0184	-0.5475	0.0184
big	-0.0374	0.7359	-0.0387	0.7269	-0.0370	0.7381	-0.0391	0.7246
sedan	-0.2552	0.5175	-0.2542	0.5190	-0.2556	0.5167	-0.2515	0.5241
channel_D	0.1393	0.1682	0.1554	0.1488	0.1383	0.1757	0.1501	0.1775
city	-0.0894	0.3427	-0.0880	0.3494	-0.0891	0.3435	-0.0887	0.3465
north	-0.2622	0.0264	-0.2629	0.0261	-0.2621	0.0265	-0.2630	0.0261
south	0.0618	0.6108	0.0624	0.6071	0.0613	0.6138	0.0616	0.6124
east	0.3787	0.1751	0.3796	0.1740	0.3793	0.1744	0.3803	0.1737
*year*2002	0.2064	0.4384	0.2047	0.4424	0.2068	0.4377	0.2050	0.4422
*year*2003	0.4234	0.0920	0.4235	0.0918	0.4249	0.0908	0.4231	0.0929
*year*2004	0.6284	0.0098	0.6272	0.0099	0.6300	0.0097	0.6286	0.0100
*year*2005	0.3021	0.2035	0.3023	0.2030	0.3037	0.2010	0.3023	0.2042
*year*2006	-0.7607	0.0014	-0.7618	0.0014	-0.7595	0.0014	-0.7617	0.0014
Adj. R-sq	0.0951		0.0949		0-0950		0-0937	

Table 8 Regression of compulsory liability claim amount in the claimed sub-sample (using the current year's maintenance record)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	10.5148	<.0001	10.5882	<.0001	10.4930	<.0001	10.5472	<.0001
vm_km_t	0.0903	0.5740					-0.3276	0.2586
vm_time_t			-0.1660	0.2314			-0.1247	0.3847
vm_item_t					0.2087	0.1548	0.4240	0.1129
*age*2530	-0.4528	0.3149	-0.4484	0.3195	-0.4467	0.3212	-0.4397	0.3288
*age*3060	-0.3945	0.3472	-0.3920	0.3501	-0.3861	0.3574	-0.3784	0.3671
*age*60*up*	-0.3790	0.4305	-0.3702	0.4411	-0.3778	0.4316	-0.3728	0.4378
female	-0.1556	0.1475	-0.1568	0.1442	-0.1603	0.1357	-0.1618	0.1321
married	0.3733	0.0343	0.3749	0.0335	0.3766	0.0327	0.3789	0.0317
*carage*1	-0.2089	0.2945	-0.1893	0.3422	-0.2197	0.2702	-0.2054	0.3039
*carage*2	-0.1895	0.3532	-0.1791	0.3801	-0.1995	0.3284	-0.1962	0.3369
*carage*3	-0.4529	0.0347	-0.4462	0.0371	-0.4644	0.0303	-0.4597	0.0321
*carage*4	-0.5508	0.0176	-0.5440	0.0190	-0.5486	0.0179	-0.5360	0.0208
big	-0.0347	0.7538	-0.0261	0.8140	-0.0350	0.7516	-0.0331	0.7659
sedan	-0.2591	0.5109	-0.2790	0.4793	-0.2648	0.5014	-0.2818	0.4747
channel_D	0.1598	0.1133	0.1027	0.3147	0.1902	0.0621	0.1480	0.1824
city	-0.0879	0.3501	-0.0923	0.3261	-0.0851	0.3654	-0.0871	0.3545
north	-0.2612	0.0270	-0.2636	0.0256	-0.2600	0.0277	-0.2622	0.0264
south	0.0599	0.6217	0.0536	0.6594	0.0578	0.6336	0.0549	0.6513
east	0.3788	0.1749	0.3767	0.1771	0.3849	0.1678	0.3894	0.1631
*year*2002	0.2002	0.4527	0.2010	0.4504	0.1854	0.4869	0.1821	0.4947
*year*2003	0.4216	0.0933	0.4275	0.0886	0.4171	0.0966	0.4220	0.0928
*year*2004	0.6242	0.0103	0.6281	0.0098	0.6105	0.0121	0.6067	0.0127
*year*2005	0.3017	0.2039	0.3019	0.2034	0.2940	0.2155	0.2886	0.2241
*year*2006	-0.7623	0.0014	-0.7605	0.0014	-0.7695	0.0012	-0.7717	0.0012
Adj. R-sq	0.0951		0.0959		0.0962		0.0963	

Table 9 Regression of compulsory liability claim amount in the claimed sub-sample (using both the previous and current years' maintenance records)

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	10.5249	<.0001	10.6000	<.0001	10.5298	<.0001	10.6025	<.0001
$vm\_km_{2\,years}$	-0.0676	0.7736					-0.0862	0.7894
$vm\_time_{2\,years}$			-0.2150	0.1540			-0.2226	0.1432
$vm\_item_{2\,years}$					-0.0356	0.8560	-0.0207	0.9389
*age*2530	-0.4504	0.3177	-0.4442	0.3239	-0.4549	0.3127	-0.4392	0.3303
*age*3060	-0.3929	0.3494	-0.3933	0.3484	-0.3974	0.3438	-0.3900	0.3535
*age*60*up*	-0.3759	0.4345	-0.3703	0.4409	-0.3810	0.4281	-0.3643	0.4492
female	-0.1537	0.1523	-0.1565	0.1447	-0.1534	0.1531	-0.1571	0.1435
married	0.3719	0.0350	0.3736	0.0341	0.3711	0.0355	0.3728	0.0347
*carage*1	-0.1990	0.3186	-0.1791	0.3695	-0.2008	0.3139	-0.1717	0.3917
*carage*2	-0.1837	0.3683	-0.1799	0.3777	-0.1846	0.3662	-0.1753	0.3911
*carage*3	-0.4420	0.0396	-0.4462	0.0371	-0.4443	0.0384	-0.4388	0.0411
*carage*4	-0.5452	0.0188	-0.5436	0.0190	-0.5467	0.0185	-0.5391	0.0202
big	-0.0386	0.7275	-0.0251	0.8213	-0.0374	0.7353	-0.0269	0.8090
sedan	-0.2553	0.5172	-0.2938	0.4567	-0.2574	0.5137	-0.2931	0.4582
channel_D	0.1387	0.1558	0.1007	0.3183	0.1398	0.1552	0.0915	0.3770
city	-0.0894	0.3417	-0.0938	0.3185	-0.0890	0.3440	-0.0951	0.3125
north	-0.2627	0.0262	-0.2592	0.0281	-0.2618	0.0267	-0.2597	0.0281
south	0.0628	0.6052	0.0562	0.6434	0.0626	0.6063	0.0571	0.6382
east	0.3774	0.1765	0.3704	0.1845	0.3794	0.1744	0.3689	0.1868
*year*2002	0.2074	0.4364	0.2027	0.4465	0.2080	0.4352	0.2049	0.4422
*year*2003	0.4244	0.0911	0.4317	0.0855	0.4259	0.0902	0.4329	0.0852
*year*2004	0.6289	0.0097	0.6313	0.0094	0.6311	0.0096	0.6338	0.0094
*year*2005	0.3025	0.2027	0.3022	0.2029	0.3044	0.2002	0.3024	0.2034
*year*2006	-0.7605	0.0014	-0.7612	0.0014	-0.7589	0.0014	-0.7609	0.0014
Adj. R-sq	0-0950		0-0962		0.0950		0.0951	

Table 10 Basic statistics of the balanced panel sub-sample

	All observations		Claimed observations	
Variables	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
claim	0.0051	0.0712		
amount			95,193	226559
vm_km_{t-1}	0.1326	0.3392	0.0855	0.2808
vm_time_{t-1}	0.4670	0.4989	0.5043	0.5021
vm_item_{t-1}	0.1637	0.3700	0.1197	0.3260
vm_km_{t}	0.1307	0.3371	0.1062	0.3095
vm_time_{t}	0.4651	0.4988	0.5310	0.5013
vm_item_{t}	0.1603	0.3669	0.1239	0.3309
$vm\_km_{2\,years}$	0.0562	0.2303	0.0171	0.1302
$vm\_time_{2\,years}$	0.2809	0.4495	0.3248	0.4703
$vm\_item_{2\,years}$	0.0809	0.2726	0.0342	0.1825
*age*2530	0.0623	0.2417	0.0256	0.1587
*age*3060	0.8989	0.3014	0.9231	0.2676
*age*60*up*	0.0332	0.1791	0.0000	0.0000
female	0.6977	0.4593	0.7692	0.4231
married	0.9356	0.2454	0.9658	0.1825
*carage*1	0.3400	0.4737	0.3500	0.4830
*carage*2	0.2405	0.4274	0.2500	0.4385
*carage*3	0.2247	0.4174	0.2100	0.4051
*carage*4	0.0010	0.0323	0.1709	0.3781
big	0.0668	0.2498	0.2308	0.4231
sedan	0.9937	0.0789	0.9829	0.1302
channel _ D	0.4720	0.4992	0.4444	0.4990
city	0.4699	0.4991	0.3590	0.4818
north	0.4565	0.4981	0.4017	0.4924
south	0.2913	0.4544	0.2906	0.4560
east	0.0236	0.1519	0.0342	0.1825
*year*2004	0.3333	0.4714	0.3333	0.4734
*year*2005	0.3333	0.4714	0.2906	0.4560
Number of observations	22977		117	

Table 11 Panel analyses of compulsory liability claim

	Model 1		Model 2		Model 3		Model 4	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Panel A: Loss probability								
Previous year								
vm_km_{t-1}	-0.5271	0.1061					-0.5151	0.3612
vm_time_{t-1}			0.1961	0.3054			0.1605	0.4119
vm_item_{t-1}					-0.4048	0.1795	0.0227	0.9638
Current year								
vm_km_t	-0.2393	0.4582					-0.2260	0.1883
vm_time_t			0.2999	0.4243			0.2625	0.5174
vm_item_t					-0.3268	0.2829	-0.4302	0.7466
Two years								
$vm\_km_{2\,years}$	-1.2702	0.0790					-0.8276	0.1898
$vm\_time_{2\,years}$			0.2737	0.1826			0.2084	0.3139
$vm\_item_{2\,years}$					-0.9526	0.0672	-0.4182	0.5483
Panel B: Loss severity								
Previous year								
vm_km_{t-1}	-0.3115	0.6099					1.1528	0.2999
vm_time_{t-1}			0.1603	0.6551			0.0448	0.1900
vm_item_{t-1}					-0.7416	0.1906	-1.5719	0.1012
Current year								
vm_km_t	0.3425	0.5821					1.2059	0.4323
vm_time_t			0.4949	0.1001			0.4848	0.1078
vm_item_t					0.1644	0.7800	-0.7289	0.6035
Two years								
$vm\_km_{2\,years}$	-0.6546	0.6022					0.0861	0.3499
$vm\_time_{2\,years}$			0.2250	0.4999			0.1765	0.6878
$vm\_item_{2\,years}$					-1.0576	0.2899	-1.3498	0.3001

Table 12 Sensitivity analyses of compulsory liability claim

		Model 1		Model 2		Model 4	
	Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Panel A: 10% threshold							
Previous year	vm_km_{t-1}	-0.0575	0.0928			-0.0593	0.1864
	vm_time_{t-1}			0.0256	0.4503	0.0222	0.5205
	vm_item_{t-1}					-0.1219	0.1319
Current year	vm_km_{t}	-0.0299	0.0313			-0.0407	0.1101
	vm_time_{t}			0.0625	0.0636	0.0783	0.0235
	vm_item_{t}					-0.0123	0.8140
Two years	$vm\_km_{2\,years}$	-0.1199	0.0663			-0.1213	0.0003
	$vm\_time_{2\,years}$			0.0476	0.2154	0.0478	0.2194
	$vm\_item_{2\,years}$					-0.1112	0.1278
Panel B: 15% threshold							
Previous year	vm_km_{t-1}	-0.0698	0.0493			-0.0756	0.1367
	vm_time_{t-1}			0.0300	0.3718	0.0307	0.3748
	vm_item_{t-1}					-0.1207	0.1340
Current year	vm_km_{t}	-0.0491	0.0201			-0.0750	0.1020
	vm_time_{t}			0.0967	0.0036	0.1237	0.0003
	vm_item_{t}					-0.0010	0.9848
Two years	$vm\_km_{2\,years}$	-0.1203	0.0433			-0.1281	0.0601
	$vm\_time_{2\,years}$			0.0691	0.0577	0.0776	0.0365
	$vm\_item_{2\,years}$					-0.1039	0.1556
Panel C: 20% threshold							
Previous year	vm_km_{t-1}	-0.0705	0.0386			-0.0678	0.1696
	vm_time_{t-1}			-0.0104	0.7007	-0.0023	0.9328
	vm_item_{t-1}					-0.0029	0.9488
Current year	vm_km_{t}	-0.0826	0.0148			-0.1036	0.0930
	vm_time_{t}			-0.0102	0.7326	-0.0263	0.3942
	vm_item_{t}					0.0161	0.7785
Two years	$vm\_km_{2\,years}$	-0.1244	0.0114			-0.1250	0.0548
	$vm\_time_{2\,years}$			-0.0153	0.6343	-0.0237	0.4654
	$vm\_item_{2\,years}$					-0.0035	0.9500

Table 13 Sensitivity analyses of compulsory liability claim using the balanced panel sub-sample

	10%		15%		20%	
Variables	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	-16.6914	0.9543	-16.6341	0.9544	-16.6703	0.9537
$vm\_km_{2\,years}$	-0.2808	0.1053	-0.3344	0.0794	-1.2702	0.0790
*age*2530	9.9782	0.9726	9.9726	0.9726	9.9319	0.9724
*age*3060	10.5576	0.9711	10.5516	0.9710	10.5003	0.9708
*age*60*up*	10.7806	0.9704	10.7720	0.9704	10.7391	0.9702
female	0.3273	0.1542	0.3257	0.1562	0.3274	0.1545
married	0.5746	0.2734	0.5737	0.2741	0.5744	0.2734
*carage*1	0.2771	0.4338	0.2721	0.4421	0.2576	0.4674
*carage*2	-0.1393	0.6831	-0.1441	0.6727	-0.1489	0.6626
*carage*3	0.2180	0.4858	0.2149	0.4920	0.2108	0.5005
*carage*4	0.1910	0.5937	0.1889	0.5977	0.1797	0.6158
big	-0.1588	0.4882	-0.1564	0.4946	-0.1583	0.4898
sedan	-0.0309	0.9666	-0.0240	0.9740	-0.0409	0.9559
channel _ D	-0.0511	0.7961	-0.0499	0.8008	-0.1126	0.5707
city	-0.4429	0.0308	-0.4426	0.0308	-0.4737	0.0209
north	-0.2745	0.2555	-0.2767	0.2513	-0.2842	0.2388
south	-0.2776	0.2824	-0.2758	0.2851	-0.2660	0.3031
east	-0.0348	0.9510	-0.0487	0.9313	-0.0386	0.9455
*year*2004	-0.1222	0.6292	-0.1166	0.6450	-0.1040	0.6813
*year*2005	-0.2612	0.2884	-0.2532	0.3037	-0.2599	0.2910
Ln Likelihood Ratio	1443.5		1442.6		1442.1	

Table 14 Two stage regression of compulsory liability claim

	Maintenance records			Claim		
Variables	**Previous year**	**Current year**	**Two years**	**Previous year**	**Current year**	**Two years**
Intercept	-0.8412***	-0.6408***	-1.2953***	-2.2778***	-2.2459***	-2.3024***
Prob(*vm _ km*)				-0.0137	-0.1186	0.1740
vm _ km				-0.0704**	-0.0823**	-0.1250**
*age*2530	-0.0958*	-0.0560	-0.0566	-0.2219**	-0.2233**	-0.2196**
*age*3060	-0.0461	-0.0007	0.0688	-0.1594*	-0.1594*	-0.1586*
*age*60*up*	0.0599	0.1038*	0.2340***	-0.1775	-0.1757	-0.1797
female	0.0184*	0.0147	0.0141	0.0714***	0.0719***	0.0709***
married	0.0195	-0.0082	0.0281	-0.0283	-0.0290	-0.0291
*carage*1	0.3578***	0.1683***	0.2198***	0.1049**	0.1058**	0.0985**
*carage*2	0.0054	0.0230	-0.0384**	0.0694	0.0702	0.0691
*carage*3	-0.0574**	-0.0333**	-0.0597**	0.0515	0.0516	0.0527
*carage*4	-0.0287	-0.0293	-0.0286	0.0350	0.0348	0.0357
big	0.0042	-0.0225**	-0.0411***	-0.0658***	-0.0667***	-0.0658***
sedan	0.0598*	-0.0019	0.0169	0.0770	0.0760	0.0758
channel _ D	-1.0504***	-1.0258***	-0.8774***	0.0161	-0.0071	0.0382
city	-0.0294***	-0.0296***	-0.0738***	-0.0799***	-0.0803***	-0.0788***
north	0.0072	0.0252**	-0.0262*	-0.2032***	-0.2026***	-0.2032***
south	0.0256**	0.0637***	0.0716***	-0.0283	-0.0262	-0.0289
east	0.1312***	0.1412***	0.1140***	0.0213	0.0250	0.0179
*year*2002	0.0074	-0.0777***	-0.0140	0.0349	0.0326	0.0355
*year*2003	0.0482**	-0.0111	0.0666**	0.0683	0.0679**	0.0682*
*year*2004	0.0249	-0.1049***	-0.0065	0.1138**	0.1102*	0.1143**
*year*2005	-0.0574***	-0.2174***	-0.1130***	0.1029**	0.0968	0.1052
*year*2006	-0.1588***	-0.1871***	-0.1553***	0.0283	0.0243	0.0320
χ^2 *value of Likelihood Ratio*	14200	13686	6501	187.39	189.66	190.37

Notes: The 99% significance level is denoted by ***.

The 95% significance level is denoted by **.

The 90% significance level is denoted by *.