Eco-friendly green drivers are indeed friendly drivers

Jerry S. Huang
Manager, Automobile Insurance Product Dept.
Fubon Insurance Co., Ltd, Taiwan

Kili C. Wang[*]
Professor, Department of Insurance,
Tamkang University, Taiwan

Abstract

Green vehicles are becoming increasingly prevalent. In following this trend, some insurance companies have provided premium discounts for these green cars. However, at first glance, the traffic accident risk of green cars seems to be higher. This paper employs a unique dataset on the mileage driven within one year for each car to formally examine whether green car drivers are friendly when driving, i.e., whether their risk of traffic accidents is lower. We use Taiwan's third-party liability insurance to test our hypotheses. We find that, before controlling for mileage driven per car, the green car drivers are not lower risk drivers. However, when we turn our attention to risk based on mileage, we instead find that the marginal effect of green cars is indeed significantly negative in terms of the claim frequency based on mileage as well as the claim severity based on mileage. Policy implications are further discussed.

[*]E-mail address: kili@mail.tku.edu.tw

Introduction

With the increasing maturity of green car technologies, the green car has become a new trend throughout the world. Many well-known insurance companies have observed this new trend, and have provided a premium discount on automobile insurance for green car owners.

Is the offer of a premium discount for green car owners by the insurance companies merely a friendly response toward the increased emphasis on environmental protection? Or it is because these green car drivers really deserve this premium discount? The answer to these questions depend on finding an answer to the question of whether the green car drivers, who are treated as eco-friendly persons, are also friendly when it comes to driving. In other words, we are interested in exploring whether the traffic accident risk is lower for these green car drivers.

The literature has argued that an individual's behavior may reflect his/her psychological or biological characteristics. Therefore, different behaviors of an individual may be correlated. For instance, the individual's driving behavior and credit score may be closely correlated (Brockett and Golden, 2007), or else the individual's driving risk and maintenance behavior may also be closely correlated (Bair et al., 2012). Hence, the green car driver's eco-friendly behavior in choosing a green vehicle may also be related to the friendly behavior when he drives. Accordingly, green car drivers may result in fewer traffic accidents and, if this is the case, they certainly deserve a premium discount.

However, some people suspect that the green car drivers could take more risks when driving than others. The reason why the green car drivers choose to purchase the green cars may not be because they are more eco-friendly, but primarily because they are high-mileage drivers. Huang et al. (2013) found that individuals who drive more tend to cause more traffic accidents involving others as well as themselves.

Hence, from this viewpoint, the green car drivers may not deserve the premium discount. The insurance companies thus have no reason to offer a special discount to these green car drivers.

Hence, we not only emphasize the importance of investigating the traffic accident risk of green car drivers, but we also point out that we should control for the mileage factor when investigating this problem. We not only investigate the traffic risk in terms of the accident frequency, but also from the point of view of the accident severity.

To investigate the above problems, Taiwan's automobile insurance market provides an ideal opportunity for such an analysis. There are two reasons for this. First, the hybrid car market is only just emerging in Taiwan, and thus far the government has not yet provided any policies to encourage the purchase of green cars. In addition, the insurance companies have also only just started the debate as to whether they should provide a green car premium discount. The individual's green car choice is not distorted by any outside forms of encouragement through policies, and fully reflects the individual's psychological and biological characteristics, as well as personal vehicle-usage considerations. In other words, Taiwan provides a neutral database to investigate this problem. Second, we are in possession of a unique dataset on the insured vehicle's mileage that is collected from the repair houses of one particular vehicle brand. By using the traffic accidents which are measured by the claim records based on the contracts of liability insurance for a sample with compulsory insurance only and a sample which also extends the coverage to voluntary insurance for one large insurance company in Taiwan, we investigate the green car drivers' comparative traffic accident risk. By merging the information regarding the vehicle's mileage, we can investigate the green car drivers' comparative traffic accident risk based on mileage.

At first, we find that the green car drivers pose a higher risk instead of being more friendly in driving, regardless of whether they are judged by frequency or severity. However, when we merge the mileage information, we find that the green car drivers are often high-mileage drivers, and their comparative traffic accident risk based on mileage is found to involve less risk than for other drivers.

These findings provide us with an important inspiration in that the green car drivers are indeed friendly in terms of their attitude toward driving. It would therefore seem that encouraging green car drivers by means of a premium discount is relevant and could improve social welfare. However, they are also high-mileage drivers, and thus a premium discount may be a misleading policy. From the point of view of the insurance companies, providing a discount on the total premium may further encourage high-mileage drivers to purchase green cars and further incentivize them to drive more because of the lower fuel expense incurred. This will result in more traffic accident losses for the insurance company. Furthermore, in spite of the green cars improving energy efficiency and lowering the amount of pollution, their higher traffic risk could give rise to an offsetting effect on social welfare. In particular, when the high-mileage drivers' motive for purchasing a green car is further driven by a premium discount, the result may be a serious loss of social welfare. So, the insurance companies should also take into account the green car drivers' mileage when providing the premium discount. When the green car drivers' lower traffic accident risk based on mileage has been proved, providing them with a discount per kilometer should be particularly relevant. This can only be implemented when the premium is priced on a basis that it is calculated according to the mileage recorded.

This paper is organized as follows. In the first section we introduce the motivation, purpose, and main findings of this research. The second section focuses on the background and hypotheses, and describes the current automobile insurance market situation, the development of a premium discount for green cars, and hypotheses which infer the traffic accident risk associated with green car drivers. The third section deals with the data, explains the structure of our research sample, and displays some preliminary observations. In the fourth section, we introduce our empirical methodologies. Section five presents the empirical results. The final section is the conclusion.

Background and Hypotheses

When we investigate whether green car drivers are less of a risk in terms of being the cause of accidents, we compare the traffic accidents caused by green cars with those caused by non-green cars. According to Taiwan's automobile insurance contract, traffic hazards are covered by the automobile's compulsory liability insurance, voluntary liability insurance, and physical insurance. However, the contract for automobile physical insurance in Taiwan suffers from serious manipulation.[1] The claims from this contract cannot exactly index the risk of a traffic accident due to the interference from such manipulation. Hence, when we explore the green cars' risk of traffic accidents, we only focus on the liability insurance.

There are compulsory as well as voluntary contracts in liability insurance. The compulsory liability insurance covers only bodily injury, and only provides a basic minimum indemnity for the third-party victim in a traffic accident regardless of whether the insured is at fault or not.[2] All of the contract's content, risk categories and premium are set by the authority. The voluntary liability insurance covers any additional bodily injury indemnity as well as the loss of property to the third-party involved in the traffic accident within the upper limit of the contract if the insured is at fault. Each insurance company has been able to design its own voluntary third-party liability insurance contract following the deregulation.

Because all drivers on the road are required to be insured by law, we can observe all the bodily injury accidents involving third parties without selection bias when we take into account the risk based on the claim filed for the compulsory liability

---

[1] Li et al. (2013) found a curious case emerging in Taiwan's automobile physical insurance contracts in that there is a significantly higher claim probability concentrated in the last policy month, which they infer as being the "premium recouping" effect. Picard and Wang (2015) also find evidence of significantly manipulative behavior in this contract.

[2] Nowadays, the indemnity for death and total disability is two million NT dollars, which is equivalent to about 67,000 US dollars.

insurance policy only. However, there are two major shortcomings. One is that we can only consider the accidents involving bodily injury. The accidents which only cause property damage are excluded from the analysis because the compulsory liability insurance indemnifies the insured against bodily injury only. The other is that the severity of the accident tends to be underestimated because the compulsory liability insurance only provides a basic minimum indemnity. To overcome these shortcomings, we ought to extend the measurement of accidents to encompass the voluntary liability insurance contract as well. However, since this extended contract is voluntarily purchased, we can only observe the accidents and their losses from those insured who have also purchased voluntary liability insurance. There is thus a sample selection problem that needs correcting when we explore the traffic accident risk on which it is based.

Accordingly, we separately investigate two sample datasets. The first sample dataset is composed of all the compulsory liability insured in our target insurance company. The traffic accidents are measured by all the claims for which the drivers are at fault.[3] The second dataset is composed of the insured who are not only covered by the compulsory liability insurance, but who have also purchased the voluntary liability insurance. Different claims, which are filed by different contracts within the same day, are counted as one accident, and the claim amounts from different contracts are summed up. In this context, we refer to such coverage as extended liability insurance for convenience. When we investigate the extended liability insurance sample, the sample selection problem is corrected in the model.

It should be asked whether the eco-friendly green car owner will also exhibit the

[3] The compulsory liability insurance also covers accidents even when the insured is not at fault. Hence, these not-at-fault accidents will not be taken into account when we perform the empirical tests.

same friendly behavior when driving. According to the literature, an individual's behavior often reflects his/her biological or psychological characteristics. This causes the behavior of different dimensions of the same person to be related to each other. In the 1990s and earlier, Lastovicka et al. (1987), Arnett (1990), Donovan (1993), and McMillen et al. (1992) pointed out that some problem behaviors, such as taking drugs, or being delinquent or hostile, could also reflect certain psychological characteristics of the individual, such as sensation, egocentrism, thrills, adventure seeking, etc. Hence, such individuals also tend to adopt risky driving habits which result in more accidents and greater losses.

By echoing this point in the above literature, Brockett and Golden (2007) and Bair et al. (2012) found the individual's driving risk to be related to his credit score or his vehicle maintenance behavior. Huang et al. (2013) also found that the individual's vehicle maintenance behavior and tendency toward fraudulent behavior are correlated. The individuals who properly maintain their vehicles are less likely to defraud. Kallunki and Pyykkö (2013) pointed out the existence of a relationship between the CEO's personal traits and his/her performance in management. The CEO's personal traits, such as overconfidence and over-optimism, are closely related to the firm's financial distress.

Following this trend of thought in the above literature, the driver's decision to purchase a green car may reflect his/her friendly trait. This friendly trait could affect his/her driving. So, we predict that the green car driver's friendly attitude has a positive effect on traffic safety. Hence, the first hypothesis measures the green car driver's friendly driving behavior in terms of accident probability/frequency such that:

Hypothesis 1: The eco-friendly green car drivers are also friendly drivers, and so their traffic accident probability/frequencies are lower than those of others.

The second hypothesis measures the green car driver's friendly driving behavior in terms of accident severity such that:

Hypothesis 2: The eco-friendly green car drivers are also friendly drivers, and so their claim amounts are lower than those of others when accidents happen and claims are filed.

However, the green car drivers could also be high-mileage drivers. Their choice of vehicle is based on considerations of lower fuel expenditures. If this is true, choosing a green car could also reflect their high mileage. According to the findings of Huang et al. (2013), high-mileage drivers are more likely to be risky drivers. Regardless of whether driving on their own or with others, the higher their average mileage, the greater their likelihood of being involved in traffic accidents. Accordingly, if they are indeed also high-mileage drivers, it is more reasonable to investigate traffic risk based on mileage instead of traffic risk without controlling for mileage when we explore the green car driver's comparative traffic risk. Hence, we first test whether the green drivers tend to be high mileage drivers by means of Hypothesis 3:

Hypothesis 3: The green car drivers tend to be high mileage drivers, which means that the green car's mileage within one year is higher than that of other cars.

When Hypothesis 3 is sustained, we will continue to test the following two hypotheses which infer traffic risk in terms of risk based on mileage:

Hypothesis 4: The eco-friendly green car drivers are also friendly drivers, and so their traffic accident probability/frequencies based on mileage are lower than those of others.

Hypothesis 5: The eco-friendly green car drivers are also friendly drivers. So, their claim amounts based on mileage are lower than those of others when accidents happen and claims are filed.

Data

We collect the individual data from the largest insurance company in Taiwan.[4] The contracts that we investigate include the compulsory liability insurance and voluntary liability insurance contracts. The information includes the characteristics of the insured and the insured vehicles which are used for underwriting and pricing, such as the gender and age of the insured, the year in which the vehicle was built, as well as the usage, brand and model. Except for the above underwriting and pricing variables, we also collect some other information including the marital status of the individual as well as the registered area of the insured vehicle.

Other than the data obtained from the insurance company, we also collect information regarding the driving mileage from the repair facilities of one particular vehicle brand manufacturer. The market share of this particular brand of vehicle is over 40%. The repair facilities provide us with information that includes the dates on which maintenance took place and the maintained vehicle's corresponding mileage record on each occasion. Hence, we can figure out each vehicle's frequency of maintenance within one year. We can also calculate the mileage recorded between two dates on which maintenance was provided. Then we can accordingly calculate the mileage driven within a period of one year for each vehicle. A list and the related definitions of these variables are provided in Table 1.

The research period covers the policy years 2011 and 2012. In other words, we investigate the policies which were underwritten during the years 2011 and 2012. Among these policies, some of the insured vehicles were purchased before March 2006, which was the year in which the green car first came to Taiwan.

Hence, we confront the sample selection problem. The insured who purchased their vehicles before this time may not have done so because they did not want to choose a

[4] The market share of this largest insurance company is over 20%.

green car, but because they had no opportunity to choose a green car. Hence, the ideal way to deal with this problem is to collect information for all insured and explain their behavior on the basis of whether they purchased a new car at this time or not. With this information to adjust for the choice of the green car, the problem of sample selection can be corrected. However, we have no information regarding the purchasing decision of those insured who purchased vehicles after March 2006. We thus do not have enough information to perform the correction. The only thing we can do is to discard all the vehicles which were purchased before March 2006. By discarding the vehicles purchased prior to March 2006, we obtain 441,920 observations in compulsory liability insurance contracts from 293,879 insured. In other words, our research sample consists of unbalanced panel data. In our research sample, 50.37% of the insured are continuously insured for two years, and 49.63% of them are insured for only one year. By focusing on the characteristics of both insured and insured vehicles, Table 2 provides an overall observation. When we focus on the sample structure of the whole compulsory liability insurance sample, we can see that the structure of the insured is very close to the structure of the insured in the whole automobile insurance market of Taiwan: more than half of the insured vehicles are registered under the name of female owners, and highly concentrated on the owners who are from 30 to 60 years old; nearly 60% of them are used in city areas, and nearly half of them are used in the northern part of Taiwan, followed by the southern, central, eastern, and outlying island parts of Taiwan. In addition, almost half of the insured vehicles are small in size, one quarter of them are medium in size, and one quarter of them are large in size. The ages of the vehicles are well distributed.

As mentioned, we may ignore the loss from the property damage and underestimate the loss from traffic accidents if we only focus on compulsory liability insurance. Hence, we further extend our research to include voluntary liability insurance. In other words, we also investigate the “extended liability insurance sample”. In this sample, there are 285,038 observations from 188,524 insured. It is also an unbalanced panel data set. 51.21% of the insured are continuously insured for two years, and 48.79% of them are insured for only one year. The structure of the characteristics of both insured and insured vehicles is also listed in Table 2. We compare the structural differences between this sample and the compulsory liability sample by means of the t test. The characteristics in terms of age and gender are similar between these two samples. The percentage of green cars is also similar between them. However, while we focus on other characteristics, compared to the compulsory liability insurance sample, there are significantly more married, more highly-educated, and more poor bonus-malus record insured in this sample.[5] The income level is comparatively lower, and the percentage of those continuing their contracts is comparatively higher in this sample. The vehicle registered area is more concentrated in the city areas and in the southern part of Taiwan in this sample than in the case of the compulsory insurance sample. The vehicle age is also comparatively more concentrated from one year to four years in this sample. In other words, there may be some sample structural differences existing between the compulsory liability insurance sample and the extended liability insurance sample.

When we turn to analyze the traffic risk based on mileage, we have to restrict our empirical analysis to a smaller group. This subgroup is made up of individuals that are insured by our target insurance company, and they are also customers in the repair facilities of our target particular vehicle brand manufacturer. When we restrict our analysis to this subgroup, the number of observations from the compulsory liability insurance sample is 19,401. Only 26% of them are continuously insured for two years. When we compare the sample structure of this subgroup with the whole group based

[5] By a highly-educated insured is meant that the individual has a PhD or master’s degree.

on the compulsory liability insurance sample using the t test, we can see that significant differences exist in all the characteristics of the insured and the insured vehicles. The percentages of green car, female, middle-aged, unmarried, highly educated, higher income and good driving record insured are significantly higher in this subgroup. The vehicles also comparatively speaking tend to be brand new, and less used in cities and the northern part of Taiwan. The insured are more concentrated in areas with higher densities of repair facilities run by the manufacturer. The number of insured who have continued their contract is also relatively high in this group.

To sum up, the sample structure of the extended liability insurance sample is quite different from that of the compulsory liability insurance sample. The significant structural difference also exists between the whole group and the sub-group which includes mileage information. Hence, if we treat the compulsory liability sample as a research base which has the least sample bias, we confront the sample selection problem when we analyze the extended liability insurance sample and also when we use the mileage information revealed by the subgroup.

When we intend to explore whether the green car drivers are more friendly in terms of their driving, there are those who may argue that the green cars are often equipped with safety equipment of a higher standard, and this may be the reason why they are metempirically safer than others if the counterpart of the green cars consists of all the other vehicles. Hence, we try to identify the influence which is purely based on the individual's behavior instead of other factors which could also affect traffic safety, such as the impact of the vehicle's safety equipment. Accordingly, as a robustness test, we will also perform tests on the other research group by comparing these green cars with their same-brand equivalent model vehicles in order to get rid of the impact of the safety equipment.

Table 3 further provides a comparison among green cars, non-green cars and the non-green cars in an equivalent model with the green cars. This comparison is performed using the compulsory liability sample, and the extended liability sample. When the comparison is performed using each sample, it is also further divided into the overall group and a sub-group with mileage information. This comparison highlights the differences by means of their claim structure and their mileage usage. Based on this preliminary observation, we find that the green cars are characterized by higher traffic risk which is measured by the claim frequency, and higher mileage which is observed by the mean level and by different percentiles. They also apparently have some different characteristics from the others by observing the sample structure in the Appendix, Table A1. As for the question of whether they are friendly drivers, this is an issue that needs to be further explored.

Methodology

When we explore the comparative traffic risk for the drivers of green cars, the variable *green* is an important variable to be examined in our empirical model. However, the problems of sample selection and endogeneity may affect this variable. The sample selection problem has already been taken care of by discarding the insured vehicles which were purchased before March 2006, as mentioned earlier.

The endogeneity problem arises in that the driver's green car choice is also one kind of decision which may affect the risk of traffic accidents. When we observe the sample structure and compare the green cars with the non-green cars,[6] we find that there is a significant difference between them. Accordingly, we can believe that some individual characteristics significantly affect the decision to choose a green car. Hence, the variable $green$ could be related to some heterogeneous factors in the regression. The endogeneity problem may exist and cause our empirical results to be biased.

When we first test Hypothesis 1, to deal with this endogeneity problem, we adopt the two-stage instrumental variables method. In the first stage, we use the instrumental variables, which are related to the choice of green cars, but are not related to the traffic accident risk, to explain the probability of choosing a green car using a Probit regression. Then, in the second stage regression, instead of directly using the dummy variable $(green)$, we use the instrumental variable $(\widehat{green})$ which is the probability of choosing a green car that is estimated in the first stage to avoid the problem of the dummy variable being correlated with the residual term.

Two factors could affect the willingness to choose a green car when an individual purchases a car. One is the individual's education level. The other is the individual's purchasing power. We first predict that an individual with a higher education level

6 The sample structures of the green car and non-green car, and a comparison of the sample structures using the t test are provided in Table A1 of Appendix 1.

may be more willing to purchase the green car. One reason is that individuals with a higher education level may be more knowledgeable about the car, and have a stronger incentive to protect the environment. They understand that the green car could be friendly to the environment, and are also more willing to be friendly to the environment by choosing a green car. The other reason is that the individuals with a higher education level may be more likely to understand how the fuel could be conserved by the green car, and this may also push them to choose a green car. However, we have no information regarding each individual's education level, and so use the percentage of the population with a higher education level (i.e., the population with a PhD or a master's degree) in the zip-code area in which the insured lives as the index. This is the first candidate for an instrumental variable (edu_i).

Secondly, the green car is also more expensive, and so the individual's wealth becomes a factor which affects the choice of a green car. While we have no real wealth level for each of the insured, we can use the average wealth level of the zip-code area in which the insured lives as the index. We use seven years of the average income level for each zip-code area to construct this index, and use it as the second candidate for an instrumental variable ($income$).[7]

Hence, the first-stage Probit regression in our instrumental variable method is:

$$Probit(green_i = 1 | edu_i, income_i, X_i)$$
$$= \Phi(\beta_{edu} edu_i + \beta_{incm} income_i + X_i \beta_X + \varepsilon_i) \qquad (1)$$

where the estimated green car choice ($\widehat{green}$) from the above regression is the instrumented variable for the dummy variable $green$.

When we explore the compulsory insurance sample further, there are no insured

[7] We use the income level from the year 2005 to the year 2011. The reasons why we choose these years are as follows. First, it makes sense that the purchasing decisions in this year could be affected by the previous year's income; second, these years are the years prior to the years in which the green car entered the market until now.

who have filed more than one claim in one contract period. Hence, we can only investigate the claim probability using the Probit regression in the second stage of our instrumental variable method:

$$Probit(clm = 1|\widehat{green}, X) = \Phi(\beta_{green}\widehat{green}_i + X_i\beta_X + \varepsilon_i) \quad (2),$$

where $clm_i = 1$ indicates that there is a claim filed by the insured during the contract period, otherwise $clm_i = 0$. X_i is a vector of explanatory variables that could affect the traffic accident hazard rate, and which includes the characteristics of the insured and insured vehicles.[8] $\hat{\beta}_{green}$ is the estimated coefficient of $\widehat{green}$ and is the key variable to be tested and observed. If the green car driver is significantly different from others in terms of traffic hazards, the coefficient should be significantly different from 0. Furthermore, if the inference from Hypothesis 1 is sustained, it should be negative. The eco-friendly green car driver is also more friendly when driving.

When we explore the extended liability insurance sample, the sample selection problem could arise in that the insured are not randomly chosen for inclusion in this sample. This is related to the insured's decision as to whether they are willing to extend their coverage to voluntary liability insurance or not. According to Heckman (1979), the sample selection problem can be corrected through a two-step method. In the first step, the selection equation is estimated. Then, the inverse Mills' ratio is calculated and included in the second step regression.

In our model, the selection equation is:

[8] In the vector of explanatory variables, we include the gender, marital status, and age of the insured. The age of the insured is further separated into three dummy variables: *age2530*, *age3060*, and *ageabv60*. The counter group contains the insured who are under 25 years old. We also include the age, size, usage, and registered area of the insured vehicle. The vehicle age is separated into *carage0*, *carage1*, *carage2* *carage3* and *carage4*. The vehicles that are more than 4 years old constitute the reference group. The registered area is separated into *north*, *south* and *central*, with east and the outlying islands as the counter group. The sizes of the vehicles are separated into *veh_m* and *veh_l*, with small vehicles serving as the reference group.

$$Probit\left(select_{extend,i} = 1 \middle| income_{\,i,} edu_{\,i}, channel_{D_i}, X_{extend,i}\right) =$$

$$\Phi(\beta_{incm} income_{\,i,} + \beta_{edu} edu_{\,i} + \beta_D channel\_D_{\,i} + X_{extend,i}\beta_{extend} + \varepsilon_i) \qquad (3)$$

The selection of whether the insured will extend his/her coverage to voluntary liability insurance is affected by the factor $X_{extend,i}$. The individual's insurance decision is often related to his/her degree of risk aversion, which can be indexed by income, as well as his/her education level. So, *income* and *edu* are included in the explanatory variables. Furthermore, the willingness to purchase voluntary automobile insurance is often related to the distribution channel. In Taiwan, car dealer ownership agents can strongly influence their customers to purchase those insurance contracts. Accordingly, *channel_D* is also included. In $X_{extend,i}$, there are other control variables which include: gender (*female*), marriage status (*married*), age (*age2530*, *age3060*, *ageabv60*), and vehicle registration area (*city*, *north*, *south*, *central*). The inverse Mills' ratio would be:

$$IMR_{extend,i} = \frac{\phi(X_{extend,i}\widehat{\beta}_{extend})}{\Phi(X_{extend,i}\widehat{\beta}_{extend})}$$

This inverse Mills' ratio will be included in the first stage of our instrumental variable method model:[9]

$$Probit(green_i = 1 | edu_i, income_i, IMR_{extend,i}, X_i)$$

$$= \Phi\,(\beta_{edu} edu_i + \beta_{incm} income_i + \lambda_{extend} IMR_{extend,i} + X_i\beta_X + \varepsilon_i) \qquad (1')$$

If the λ_{extend} in the above regression is significantly different from 0, this means that there is a sample selection problem. This can be corrected by including this inverse Mills' ratio, and the correct instrumented variable ($\widehat{green}_i$) is accordingly measured. Furthermore, some insured in this extended liability insurance sample have not only filed a claim, but they have also filed more than one claim at different times.

[99] We also use the error correction method raised by Murphy and Topel (1985) to correct the estimated asymptotic covariance matrix when we use the inverse Mills' ratio. In all of our following Probit regressions, when we correct the selection problem by using the inverse Mills' ratio, we also perform this error correction in all of them.

Hence, we investigate their traffic accident frequency. When we use negative binomial regression to evaluate the accident frequency, apart from the endogeneity problem, the selection problem also exists. In addition to including the instrumented variable ($\widehat{green_i}$) in the explanatory variables, we also correct the selection problem by including the inverse Mills' ratio:[10]

$$number_i \sim Poisson(\mu_i)$$

$$\mu_i = \exp(\beta_{green}\widehat{green_i} + \lambda_{extend}IMR_{extend,i} + X_i\beta_X + \varepsilon_i) \quad (4)$$

where $number_i$ is the number of claims that the insured filed within the policy year, and ranges from 0 to n. The estimated coefficient ($\hat{\beta}_{green}$) is still the key coefficient to be tested and observed. If the inference from Hypothesis 1 that green car drivers are eco-friendly is sustained, it should be negative and significantly different from 0.

In Hypothesis 2, we are interested in comparing the severity of traffic accidents between green cars and non-green cars. When we investigate this hypothesis based on compulsory liability insurance, the "green" variable is the key to be investigated, and the endogeneity problem also needs to be taken care of. When we use OLS regression to analyze the claim severity, we adopt the treatment effect model raised by Maddala (1983):

$$clmamt_i = \beta_{green}green_i + X_i\beta_X + \varepsilon_i \quad (5)$$

$$green_i^* = \gamma_0 + X_{treat}\gamma_{treat} + u_i$$

[10] We as well use Murphy and Topel's the error correction to correct the estimated asymptotic covariance matrix when we use the inverse Mills' ratio. Although, Terza (1995) argue that the selected subpopulation would surely not apply a same distribution as the original model. Hence, when he argued on the Poisson regression model. He believed that the connection between the selection model and the Poisson conditional mean function is ill defined. Hence, he suggest another approach by means of the conditional negative binomial regression. He directly incorporated the selection mechanism in a negative binomial regression by a conditional function, and jointly evaluated them by the joint normality of the two error terms from the selection equation and negative binomial regression. However, Greene (1995) has proven that Teza's model reduce to Green's approach. In words, adjusted by the error correction method raised by Murphy and Topel (1985) can achieve a very close results.

$$green_i = \begin{cases} 1, & green^* > 0 \\ 0, & otherwise \end{cases}$$

If we investigate his hypothesis by incorporating extended liability insurance, the sample selection problem also needs to be taken care of by including the inverse Mills' ratio (IMR_{extend}; which is measured by including regression (3)) in the above treatment effect model:

$$clmamt_i = \beta_{green} green_i + \lambda_{extend} IMR_{extend,i} + X_i \beta_X + \varepsilon_i \qquad (5')$$

$$green_i^* = \gamma_0 + \lambda_{extend} IMR_{extend,i} + X_{treat} \gamma_{treat} + u_i$$

$$green_i = \begin{cases} 1, & green^* > 0 \\ 0, & otherwise \end{cases}$$

The variable $clmamt_i$ is the claim amount per accident. It is measured in thousands of NT dollars. The key variable used to judge the relatively high traffic accident severity from green cars is the estimated coefficient $\hat{\beta}_{green}$. If it is positive and significantly different from 0, it means that the green cars face a more severe loss when an accident occurs, and our Hypothesis 2 is sustained.

As we have mentioned, the green car drivers may also be high mileage drivers. Hence, we further test whether the green car drivers tend to drive more in accordance with Hypothesis 3 by using OLS regression. However, we have to restrict our research to a smaller subgroup which is made up of individuals that are insured by our target insurance company, and they are also customers in the repair facilities of our target particular vehicle brand manufacturer. However, when we compare the whole of the compulsory liability insurance sample with this subgroup in Table 2, a significant sample structural difference is found to exist. Concern with the sample selection problem accordingly emerges. We thus need to correct the sample selection problem according to the following selection equation:

$$Probit\left(select_{repair,i} = 1 \middle| income_{i,} edu_i, access_i, X_{repair,i}\right) =$$

$$\Phi(\beta_{incm} income_{i,} + \beta_{edu} edu_i + \beta_{acs} access_i X_{repair,i} \beta_{repair} + \upsilon_i) \qquad (6)$$

where, X_{repair} includes the factors which would affect the decision of the insured as to whether to repair their vehicles in the repair facilities of the manufacturer or not. As observed from Table 2, all the structures of the characteristics of the insured and the insured vehicles are significantly different between the subgroup and the group as a whole. The structures of education level and income level are also significantly different between them. Therefore, these factors should be included in X_{repair}. Apart from these, whether it is convenient to maintain or repair a vehicle in the repair facilities of the manufacturer could affect the insured's decision when choosing a repair location. Hence, the number of the manufacturer's repair facilities in the zip-code area in which the insured lives (*access*) is also one potential factor to be included.

The inverse Mills' ratio is included in the analysis in the following treatment effect model to correct the sample selection problem:

$$mileage_i = \beta_{green} green_i + \lambda_{repair} IMR_{repair,i} + X_i\beta_x + \varepsilon_i \qquad (7)$$

$$green_i^* = \gamma_0 + \lambda_{repair} IMR_{repair,i} + X_{treat}\gamma_{treat} + u_i$$

$$green_i = \begin{cases} 1, & green^* > 0 \\ 0, & otherwise \end{cases}$$

When we investigate the comparative mileage driven for green cars based on the compulsory insurance sample, IMR_i includes $IMR_{repair,i}$ only. It is calculated from regression (6). When we investigate the comparative mileage driven for green cars based on extended liability insurance, IMR_i includes not only $IMR_{repair,i}$, but also includes $IMR_{extend,i}$. $mileage_i$ is the insured's mileage within one year, and it is measured by thousands of kilometers driven. If the estimated coefficient $\hat{\beta}_{green}$ is positive and significantly different from 0, which means that the green car drivers tend to drive more, Hypothesis 3 is supported.

If the green car drivers are accustomed to high mileage, this could be the reason for the increase in their traffic risk. In this case, we then control for the mileage when we explore whether the green car drivers are friendly drivers. We now turn to investigate the comparative accident frequency based on mileage within one year for green cars, instead of merely investigating the comparative accident probability/ frequency per year for green cars according to Hypothesis 4. We also instrument the variable $green$ at the first stage. Then, at the second stage, we investigate the number of claims per thousand kilometers within one year using the Tobit regression. When we analyze the compulsory liability insurance sample, there is still a sample selection problem raised by the observations which are included in the subgroup with mileage information. When we analyze the extended liability insurance sample, apart from the above sample selection problem, the other sample selection problem will also be raised when the insured who are included in this sample are not randomly chosen. Hence, the inverse Mills' ratio (IMR_i) should be included in the instrumental regression and the following Tobit model:

$$number_{per_i}^{*} = \beta_{green} green_i + IMR_i \lambda + X_i \beta_X + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

$$number_{per_i} = \begin{cases} number_{per_i}^{*} & if \;\; number_{per_i}^{*} > 0 \\ 0 & if \;\; number_{per_i}^{*} < 0 \end{cases} . \tag{8}$$

IMR_i is IMR_{repair} when we analyze the compulsory liability insurance sample, and IMR_i includes IMR_{repair} and IMR_{extend} when we analyze the extended liability insurance sample. $number_{per_i}$ is the number of claims per thousand kilometers, and equals $number_i / mileage_i$. We focus on the estimated coefficient $\hat{\beta}_{green}$. When $\hat{\beta}_{green}$ is negative and significantly different from 0, it means that the green car drivers have fewer claims per thousand kilometers than the other drivers. The green car drivers are safer than the other drivers according to the accident frequency based on mileage. This is evidence that supports Hypothesis 4.

In accordance with Hypothesis 5, we also further explore the comparative accident severity based on mileage for the green car drivers using the treatment effect model:

$$clmamt\_per_i = \beta_{green} green_i + IMR_i\lambda + X_i\beta_X + \varepsilon_i \qquad (9)$$

$$green_i^* = \gamma_0 + IMR_i\lambda + X_{treat}\gamma_{treat} + u_i$$

$$green_i = \begin{cases} 1, & green^* > 0 \\ 0, & otherwise \end{cases}$$

In order to correct the sample selection problem, one inverse Mills' ratio (IMR_{repair}) is also included in the regressions when we analyze the compulsory liability insurance sample, and two inverse Mills' ratios (IMR_{repair} and IMR_{extend}) are also included in the regressions when we analyze the extended liability insurance sample. $clmamt_{per_i}$ represents the claim amount per thousand kilometers, and equals $clmamt_i/mileage_i$. If $\hat{\beta}_{green}$ is negative and significantly different from 0, this means that the green car drivers have fewer severe accidents than other drivers when compared based on the criterion of the claim amount per thousand kilometers. Our Hypothesis 5 is thus sustained.

Empirical results

We preliminarily observe from the compulsory liability insurance that the green car drivers seem to have higher probabilities of claims, as well as greater numbers of claims. Hence, we use a Probit regression to perform a formal test. Apart from the claim probability, the severity when they claim is also explored using OLS regression. [11]

Table 4 displays the empirical results. When we use Probit regression to test the green cars' comparative claim probability, in order to deal with the possible bias caused by the endogeneity problem, we adopt the two-stage instrumental variables method. In the first stage, we use two instrumental variables, $income$ and edu to instrument the variable $green$. Based on the Durbin-Wu-Hausman test, the null hypothesis of no endogeneity is rejected at the 1% significance level, so that we can confirm that the instrumental variable method is relevant. Based on the Anderson-Rubin test, the null hypothesis whereby the instrumental variable is exogenous in this instrumental regression cannot be rejected, so that we can confirm that these two variables are not weak instrumental variables. Based on Sargan's J test, the null hypothesis of no over-identification cannot be rejected either, so that we can confirm that our instrumental variable is not over-identified. Overall, the above tests confirm that our instrumental variable method is appropriate.

We can also observe that the variables of *income* and *edu* have significant explanatory power in explaining the choice of green cars. The estimated coefficient of *income* is significantly different from 0 at the 1% significance level, and the estimated coefficient of *edu* is significantly different from 0 at the 5% significance level. The estimated coefficient $\hat{\beta}_{incm}$ is positive, meaning that people with higher purchasing

[11] Our research samples consist of unbalanced panel data. Hence, the regressions in this paper are performed using the random effects model.

power, which is indexed by the wealthier areas in which they live, are more willing to purchase a green car. The education level of the insured is also an important factor in purchasing a green car since the estimated coefficient $\hat{\beta}_{edu}$ is positive, which means that people with a higher education level will have a greater tendency to choose a green car.

The by-product is that we also find that the insured who are within or above the middle age range, live in city areas, and have a better bonus-malus record will be more likely to purchase green cars.

In the second stage, we can observe that the estimated coefficient of the instrumented variable $green$ is positive, but it is not significantly different from 0. Our Hypothesis 1 thus cannot be sustained based on these empirical results.

We also investigate the comparative severity of loss from green cars when traffic accidents occur. This is performed by the treatment effect model, and the outcomes are also displayed in Table 4. We find that the individual's income level and education level still have a significantly positive effect in terms of purchasing a green car. The by-product is that those individuals who have better driving records, as observed by their bonus-malus coefficients, have a greater tendency to purchase green cars.

In the regression used to estimate the claim amount, the λ coefficient of the inverse Mills' ratio is significantly different from 0. The Hausman test is also is shown to reject the null hypothesis of "no endogeneity problem" at the 1% significance level. Hence, we can confirm the existence of the endogeneity problem, and performing the test using the treatment effect model is appropriate. We also find that the estimated coefficient of the variable $green$ is positive, but it is not significantly different from 0. Our Hypothesis 2 cannot be sustained based on these empirical results either.

We also perform the same test using the two-stage instrumental variables method and the treatment effect model for the green cars' comparative claim probability and comparative claim severity by taking the equivalent model vehicles as the counterpart. This is a robustness test to identify the decision behavior of the green car owners from the effect of the safety equipment. The empirical results are presented in the same table. We find a consistent outcome that the green cars' claim probability and claim severity are not lower than for the others. They are higher, although the empirical evidence is not statistically significant. The endogeneity problems are also proven to exist in this robustness test.

We then turn to test these empirical works in the extended liability insurance sample. The sample selection problem emerges by using this sample. Hence, in comparison with the empirical results in Table 4, one more thing has to be taken care of. We correct the sample selection bias by including the inverse Mills' ratio estimated from a selection equation of the decision as to whether to extend the coverage to voluntary liability insurance or not. The results of the selection equation are listed in Appendix 2. We find that individuals' income level, education level, and whether or not they purchase the insurance contract from car dealers have a strong effect on this selection. Higher income individuals are less likely to purchase extended contracts, while more highly educated individuals and the ones who purchase the insurance through car dealers are more likely to purchase extended contracts. The by-product is that females, middle-aged persons, and those who live in cities have a stronger tendency to purchase extended insurance.

By the inverse Mills' ratio calculated from this equation, we correct the sample selection problem in the analyses of the two-stage instrumental variables method and treatment effect model. The sample selection bias is proved to exist by observing the estimated λs of the IMR_{extend} in the regressions in Table 5. All of them are significantly different from 0.

The other thing which is worth mentioning is that we test the comparative claim frequencies for green cars in this sample instead of measuring the comparative claim probability, because there are many insured who have filed a claim on more than one occasion in this sample. When the test is performed for the extended liability insurance sample, we find that the green cars are shown to exhibit a higher claim frequency. The estimated coefficient of the instrumented *green* is positive and different from 0 at the 5% significance level. The green cars also face a more severe loss when the claims are filed. The estimated coefficient of *green* in the second stage of the treatment effect model is positive and different from 0 at the 10% significance level. These results are also robust when we perform the robustness test by comparing the green cars with their equivalent model vehicles.[12] In other words, Hypothesis 1 and Hypothesis 2 still cannot be sustained by the empirical tests when we use the extended liability insurance sample.

However, before we cease to believe that the green car drivers are friendly drivers, we should figure out whether their higher comparative accident probability/frequency and accident severity occur because they drive more. Hence, we now turn to test Hypothesis 3.

When we include the mileage information in our analysis, our research sample is further restricted to a smaller subsample which is formed by merging the data from our target insurance company and the repair house which is owned by one particular brand of vehicle. The other sample selection problem rises accordingly. The individuals' decision whether to repair their vehicles in the manufacturer's repair facilities is measured. In addition, the empirical results for the selection equation when we use the whole compulsory liability insurance sample, and the selection

[12] When the robustness test based on an equivalent sample is performed using the extended liability insurance sample, the selection bias should also be taken care of. The selection equations are listed in Appendix 2.

equation when we use the equivalent sample are provided in Appendix 2. We can see that the convenience of access to the manufacturer's repair facilities, as well as the income level and education level of the individuals together have a significant impact on the decision as to whether they will have their vehicles repaired in these repair houses. Those individuals who live in areas which have more of the manufacturer's repair houses, those with higher income levels and those with higher education levels are more willing to repair their cars in these repair houses.

We also observe that females, older individuals and new car owners are more willing to have their vehicles repaired by the manufacturer. Those owners who live in cities or in the northern part of Taiwan are less likely to go to the manufacturer's repair house.

By correcting the sample selection problem through the inverse Mills' ratio calculated from the above equations, we can test the comparative mileage usage of green cars using the treatment effect model. The empirical results are summarized in Table 6. The finding is that the green cars are indeed driven more than other vehicles. The estimated coefficient for *green* in the second stage regression is positive and significantly different from 0. The robustness test whereby the green cars are compared with their equivalent model vehicle counterparts gives rise to similar results.

We also perform these tests for Hypothesis 3 based on the extended liability insurance sample in Table 7. One more sample selection problem, involving the decision of whether to extend the insurance to voluntary liability insurance or not, is also taken care of. Hence, there are two inverse Mills' ratios (IMR_{repair} and IMR_{extend}) in the empirical model.[13] The results of the robustness test to compare

[13] The empirical results from these selection equations are basically consistent with those listed in Table 2A. Because there are many selection equations that have to be run, displaying them one by one

green cars with an equivalent model vehicle counterpart are also presented in this table. All the results indicate that the green cars are driven more than the other cars. The inference of Hypothesis 3 is thus sustained.

Since the evidence shows that the green cars are driven more, we can compare their accident frequency and relative claim amounts based on mileage with those for non-green cars. In other words, we test Hypothesis 4 and Hypothesis 5 in what follows.

The empirical results, which are tested using the compulsory liability insurance sample, are listed in Table 8. The sample selection problem also arises because we restrict the analysis to the subgroup which includes the mileage information. This problem has to be corrected using the inverse Mills' ratio (IMR_{repair}) in both the first and second stage regressions. From observing the λs of these inverse Mills' ratios, it can be concluded that all of them are significantly different from 0, which means that there is indeed a sample selection problem and that this problem needs to be corrected.

The claim frequencies per thousand kilometers are tested by using the two-stage instrumental variable Tobit model. The variable $green_i$ is instrumented in the first stage as well. The J test, Anderson-Rubin test, and Durbin-Wu-Hausman test help us to confirm that our instrumental variable method is relevant. Both $\hat{\beta}_{incm}$ and $\hat{\beta}_{edu}$ are positive and significantly different from 0. The insured with higher purchasing power and higher education levels also tend to purchase green cars. The by-product observation is also consistent with our previous results. The estimated coefficient of the instrumented variable $green_i$ is negative and significantly different from 0 at the 1% significance level. It implies that, when observing the accident frequency per

would be redundant, and thus we do not list them one by one. These results can, however, be provided upon request.

thousand kilometers, the green cars are significantly less risky than the others. This evidence supports the inference of our Hypothesis 4.

The above outcome is also confirmed by the robustness test when we compare the green cars with their equivalent model vehicle counterparts. Although the statistical significance level and the values of the estimated coefficient of this instrumented variable $green_i$ are both lower than those for the original whole sample, the estimated coefficient is still negative and significantly different from 0 at the 5% significance level. Hence, the concern that the vehicle's safety equipment factor makes the green cars metempirically safer than other vehicles does not affect our inference that the green car drivers are friendly on the road. The inference of Hypothesis 4 is thus sustained by these empirical results.

The claim amounts per thousand kilometers are tested by the treatment effect model. As the results of the first stage regression indicate, both the $\hat{\beta}_{incm}$ and the $\hat{\beta}_{edu}$ are positive and significantly different from 0. The inference related to the influence of the individual's income level and education level remains as well. The main empirical result in the second stage is that the estimated coefficient of $green_i$ is negative and different from 0 at the 1% significance level. The robustness test based on comparing the green cars with the equivalent model vehicles results in the same outcomes. The inference of Hypothesis 5 is thus sustained, and it is also not affected by concerns over the safety equipment in the vehicles.

The empirical results, which are obtained using the extended liability insurance sample, are listed in Table 9. The sample selection problem also arises not only because we restrict the analysis to the subgroup, but also because the insured in this sample are not randomly chosen. This problem has to be corrected by two inverse Mills' ratios (IMR_{repair} and IMR_{extend}) in both the first and second stage regressions. The two sample selection problems are also shown to exist, and so correcting them is necessary.

We still use the instrumental variable Tobit model to estimate the comparative claim frequencies for green cars. This instrumental variable model is also proved to be appropriate based on the results of the Hausman test, J test and AR test. The estimated coefficient of the instrumental variable $green_i$ is negative and significantly different from 0 at the 1% significance level just as the results from the compulsory liability insurance sample indicate. This finding also implies that the green cars are significantly less risky than the others based on the judgement of accident frequency per thousand kilometers. The inference of Hypothesis 4 is supported as well. The results of the robustness test when comparing the green cars with equivalent model vehicles in Table 9 also support Hypothesis 4.

We also use the treatment effect model to test the comparative claim severity for green cars in this sample. The factors which affect the decision to purchase green cars have the same effect as the results for the compulsory liability insurance sample. The estimated coefficient of $green_i$ is also negative and different from 0 at the 1% significance level. The losses on green cars are significantly less severe when we evaluate risk severity based on mileage. The inference of Hypothesis 5 is thus also sustained.

The robustness test also displays consistent empirical outcomes. The inference of Hypothesis 5 is sustained again under the empirical tests conducted in this sample.

So far, the inference has been that the green car drivers are friendly drivers in accordance with the criterion of claim frequencies based on mileage. The inference is also supported by the criterion of the claim amount based on mileage. Furthermore, this inference is fully sustained by different research samples and the robustness test.

After performing all the empirical tests, it is found that although the green car drivers do not seem to be friendly drivers at first sight, green car drivers are found to be less risky when we analyze their claim frequency per thousand kilometers and the claim amounts per thousand kilometers. All of the above results confirm that in spite of the green cars' overall higher accident risk, the risk of an accident that they pose based on mileage is still lower than for non-green cars.

Our empirical outcomes not only infer that the green car drivers are friendly drivers, but also provide important policy implications. Overall, the "appear more risky" attribute of the green car drivers mainly arises due to the fact that they drive more. The green car drivers are indeed friendly drivers, and thus they deserve to receive a premium discount. However, we suggest that the premium discount should be provided according to the mileage driven. In other words, if the insurance companies can calculate the premium based on the mileage driven, providing a discount for the green drivers based on the mileage driven is relevant.

We can also briefly observe some economic effects in our analysis. If we provide a premium discount for green car drivers without accounting for their mileage, the consequence can be inferred by our empirical results in Tables 4 and 5. Based on the instrumental variables method in Table 5,[14] the estimated coefficient of $green$ based on the whole sample is 0.1233, and its IRR is calculated as 1.1097. This means that the accident frequency is increased by 10.97% more for green cars than for the other cars. According to the robustness test, the estimated coefficient of $green$ is 0.0897, and its IRR is calculated as 1.0763. This means that the accident frequency is increased by 7.63% more for green cars than for the equivalent model vehicles.

Based on the treatment effect model shown in Table 5, the estimated coefficient of $green$ for the whole sample is 54.7052, which means that the green cars have a claim cost per accident that exceeds that for the non-green cars by NT$ 54,705. Based

[14] We only roughly calculate the economic effect in Table 5 instead of Table 4 because the estimated coefficients of *green* in Table 4 are not statistically significantly different from 0. Hence, the following calculations regarding the economic effect also only focus on the extended liability insurance sample.

on the robustness test, although the estimated coefficient of $green$ is 30.3832, it is not significantly different from 0. Overall, at first glance the results indicate that the green car drivers have a higher expected claim cost than the non-green drivers.

However, based on our empirical results, we are able to find out whether the green car drivers are accustomed to being high mileage drivers. Based on the same sample, in Table 7, we find that the green cars are driven 7,081 kilometers more on average than the other cars within a period of one year.[15] According to the robustness test sample, the green cars are driven 5,051 kilometers more on average than their equivalent model vehicles.[16] When higher mileage induces a higher accident frequency and more severe accidents,[17] we ought to measure the accident frequency per thousand kilometers and the claim amount per thousand kilometers. We can also observe the economic benefits accruing due to the green drivers' lower accident cost per thousand kilometers as follows.

The accident frequency per thousand kilometers is observed from Table 9. The estimated coefficient in the Tobit regression represents the marginal effect of the explanatory variables on the explained variable for those insured who have filed claims. In the extended sample, the estimated coefficient of $green_i$ is -0.5109. This implies, in terms of the thousands of kilometers driven, that the green car drivers will make 0.5109 fewer claims than others. The marginal effect of the explanatory variables on the explained variable for all insured, regardless of whether they file a claim or not, is calculated by the estimated coefficient in the Tobit regression

[15] It is 19.3989 kilometers more on average each day.
[16] It is 13.8371 kilometers more on average each day.
[17] In our two research samples, we investigate the increments in accident frequency and accident severity for each additional thousand kilometers driven. Based on the negative binomial regression, when the insured drive one thousand kilometers more, the claim frequency will increase by 0.1896 times in the equivalent sample, and the claim frequency will increase by 0.1341 times in the matched sample. Based on the OLS regression, when the insured in the equivalent sample drive an additional thousand kilometers, the claim amount will increase by NT$ 1,045 when they file the claim; when the insured in the matched sample drive an additional thousand kilometers, the amount of the claim will increase by NT$ 831 when they file the claim.

multiplied by its mean value. In this sample, in terms of thousands of kilometers driven, the expected claim frequency of green car drivers is 0.0895 less than for others while the claim probability is 0.1751. The amount of the claim for each thousand kilometers driven is analyzed in the same table. In terms of thousands of kilometers driven, the amount claimed is NT$1,700.5 less for green car drivers than for others.

In the robustness test in this same table, the estimated coefficient of $green_i$ is -0.3278. It implies, in terms of thousands of kilometers driven, that the green car drivers will make 0.3278 fewer claims than others. When the claim probability is 0.1417, the expected claim frequency of green car drivers is 0.0464 less than for others in terms of thousands of kilometers driven. The amount of the claim for each thousand kilometers driven is also in the same table. In terms of thousands of kilometers driven, the amount claimed is NT$1,150.8 less for green car drivers than for others.

Conclusions

With green cars becoming more and more prevalent nowadays, the debate as to whether the green car drivers deserve to enjoy a discount on their premiums when they purchase automobile insurance has only just begun. Do the green car drivers really deserve this discount? The answer to this depends on whether the green car drivers are lower risk drivers or not.

By using automobile insurance data from one large insurance company in Taiwan, we explore the traffic accidents that occurred according to each of the compulsory and voluntary liability insurance contracts and automobile physical insurance contracts.

At first glance, we find that the green car drivers face relatively higher traffic risk than others. However, we also find that the green car drivers may choose green cars because they are high mileage drivers. They tend to drive more. Both the empirical literature and our research sample have proved that higher mileage drivers often face a greater risk of traffic accidents. Therefore, this may be the reason why we empirically find that the green car drivers face a higher degree of risk in terms of both accident frequency and accident severity.

So, we further analyze the drivers of green cars versus non-green cars based on accident frequency per thousand kilometers and amounts claimed per thousand kilometers. We find that the accident frequency per thousand kilometers for green car drivers is lower than for other car drivers. The accident severity per thousand kilometers of green car drivers is also lower than that for other drivers. Both of the inferences from Hypothesis 4 and Hypothesis 5 are sustained by our empirical results.

These findings provide practitioners with an important policy implication in that the green drivers are indeed friendly drivers. They should thus deserve a premium discount. However, they tend to drive more, and thus their overall risk is accordingly higher. If the insurance companies are not aware of this and simply provide a discount for the green car drivers, this discount strategy will cause the insurance companies to lose money due to the higher expected cost of insuring the green cars. The green cars have between 0.1097 and 0.0763 higher claim frequencies than the non-green cars. Their claim amounts per accident range from NT$ 54,705 to NT$30,383 more than for non-green cars.

Accordingly, the insurance companies should provide a premium discount based on the mileage driven. In other words, if the insurance company can calculate the premium based on the mileage driven, it is relevant to provide a discount based on mileage instead of simply providing a premium discount for green car drivers that is not based on a premium that is calculated according to the mileage driven.

References:

Arnett, J., 1990, Drunk driving, sensation, and egocentrism among adolescents, ***Personality and Individual Difference***s, 11(6): 541-546.

Bair, S. N, Huang, R. J., and Wang, K. C., 2012, Can vehicle maintenance records predict automobile accidents? ***Journal of Risk and Insurance***, 79: 567-584.

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, ***The Journal of Risk and Insurance***, 74:1; 23-63.

Donovan, J. E., 1993, Young adult drinking-driving: Behavioral and psychosocial correlates, ***Journal of Studies on Alcohol***, September: 600-613.

Greene, W. H., 1995, Sample selection in the Poisson regression model, NYU ***working paper***, No. 2451/26256, New York University, Stern School of Business.

Heckman, J., 1979, Sample selection bias as a specification error, ***Econometrica*** 47: 153-161.

Huang, R. J., Tzeng, L. Y., and Wang, K. C., 2013, Heterogeneity of accident externality from driving. ***Journal of Risk and Insurance***, DOI: 10.1111/j.1539-6975.2013.01517.x**.**

Kallunki, J. P. and Pyykkö, E., 2013, Do defaulting CEOs and directors increase the likelihood of financial distress of the firm? ***Review of Accounting Studies***, 18 (1): 228-260.

Lastovicka, J. L., Murry, J. P., Joachimsthaler, E. A., Bhalla, G. and Scheurich, J., 1987, A lifestyle typology to model young male drinking and driving. ***The Journal of Consumer Research***, 14(2): 257-263.

Li C. S., Liu, C. C. and Peng, S. C., 2013, Expiration dates in automobile insurance contracts: the curious case of last policy month claims in Taiwan, ***The Geneva***

Risk and Insurance Review, 38 (1): 23-47.

Murphy, K. and Topel, R., 1985, Estimation and inference I two steps econometric models, ***Journal of Business and Economic Statistics***, 3: 370-379.

Maddala, G. S. 1983, Limited-dependent and qualitative variables in econometrics, ***Cambridge University Press***.

McMillen, D. L., Pang, M. G., Wells-Parker, E. and Anderson, B. J., 1992, Alcohol, personality traits, and high risk driving: A comparison of young, drinking driver groups, ***Addictive Behaviors***, 17: 525-532.

Picard, P. and Wang K. C., 2015, Insurance fraud through collusion between policyholders and car dealers: Theory and empirical evidence, ***working paper***.

Terza, J., 1995, Estimating count dada models with endogenous switching and sample selection, ***working paper***, Department of Economics, Pennsylvania State University, State College, PA, March.

Table 1 Variable definitions

Variables	Definition
number	The number of claims filed by the insured
clm	A dummy variable that equals 1 if the insured has already filed a claim, and 0 otherwise.
clmamt	A variable that represents the total claim amount in each traffic accident, measured in thousands of NT dollars.
mileage	A variable that represents the mileage recorded by the insured car in one year, measured in thousands of kilometers.
green	A dummy variable that equals 1 if the insured vehicle is a green car, and 0 otherwise.
access	A variable that implies the accessibility of the green car. It is represented by the number of sales outlets and repair shops of the green car in the zip-code area in which the insured lives.
income	The average income level of each zip-code area in which the insured lives. The average income level is calculated based on the seven years of income from the year 2005 to the year 2011.
edu	The percentage of the population with a higher education level for each zip-code area in which the insured lives. The definition of "higher education level" is that the person has either a PhD or a master's degree.
channel_D	A dummy variable that equals 1 if the insured has purchased the insurance contract through the car dealer-owned agent, and 0 otherwise.
female	A dummy variable that equals 1 if the insured is female, and 0 otherwise.
married	A dummy variable that equals 1 if the insured is married, and 0 otherwise.
age2025	A dummy variable that equals 1 if the insured is between the ages of 20 and 25, and 0 otherwise.
age2530	A dummy variable that equals 1 if the insured is between the ages of 25 and 30, and 0 otherwise.
age3060	A dummy variable that equals 1 if the insured is between the ages of 30 and 60, and 0 otherwise.
ageabv60	A dummy variable that equals 1 if the insured is above 60 years old, and 0 otherwise.
city	A dummy variable that equals 1 when the owner of the car lives in a city, and 0 otherwise.

north	A dummy variable that equals 1 when the car is registered in the north of Taiwan, and 0 otherwise.
south	A dummy variable that equals 1 when the car is registered in the south of Taiwan, and 0 otherwise.
central	A dummy variable that equals 1 when the car is registered in the central part of Taiwan, and 0 otherwise.
carage0	A dummy variable that equals 1 when the car is brand new, and 0 otherwise.
carage1	A dummy variable that equals 1 when the car is one year old, and 0 otherwise.
carage2	A dummy variable that equals 1 when the car is two years old, and 0 otherwise.
carage3	A dummy variable that equals 1 when the car is three years old, and 0 otherwise.
carage4	A dummy variable that equals 1 when the car is four years old, and 0 otherwise.
veh_m	A dummy variable that equals 1 when the insured car has an engine capacity that is above 1800 c.c. and equal to or less than 2000 c.c., and 0 otherwise.
veh_l	A dummy variable that equals 1 when the insured car has an engine capacity equal to or greater than 2000 c.c., and 0 otherwise.
sedan	A dummy variable that equals 1 when the car is a sedan and is for non-commercial or for long-term rental purposes, and 0 otherwise.

Table 2 Sample structure

	Compulsory (A)	Extended (B)	Subgroup (C)	(A)-(B)	(A)-(C)
green	0.0141	0.0138	0.0374	0.0003	-0.0233***
female	0.6153	0.6159	0.7327	-0.0006	-0.1175***
married	0.5621	0.5700	0.1432	-0.0078***	0.4190***
age2025	0.0141	0.0139	0.0095	0.0002	0.0045***
age2530	0.0551	0.0544	0.0580	0.0007	-0.0028*
age3060	0.8424	0.8429	0.8682	-0.0004	-0.0258***
ageabv60	0.0878	0.0882	0.0639	-0.0004	0.0239***
city	0.6180	0.6183	0.5631	-0.0003**	0.0549***
north	0.4969	0.4921	0.3773	0.0049***	0.1197***
south	0.2695	0.2726	0.3390	-0.0031***	-0.0695***
central	0.1802	0.1815	0.2271	-0.0013	-0.0468***
carage0	0.1608	0.1417	0.9680	0.0191***	-0.8072***
carage1	0.1608	0.1639	0.0315	-0.0031***	0.1294***
carage2	0.1434	0.1471	0.0005	-0.0037***	0.1429***
carage3	0.1377	0.1402	0.0001	-0.0025***	0.1377***
carage4	0.1283	0.1315	0.0000	-0.0032***	0.1283***
veh_m	0.2553	0.2550	0.1842	0.0004	0.0712***
veh_l	0.2335	0.2342	0.1725	-0.0007	0.0610***
sedan	0.8812	0.8809	0.9686	0.0003	-0.0875***
bm	0.7444	0.7422	0.8147	0.0022***	-0.0703***
income	27.2775	27.2362	28.8210	0.0413**	-1.5435***
edu	0.0683	0.0684	0.0745	-0.0001***	-0.0062***
access	1.8592	1.8581	1.9577	0.0011	-0.0985***
cont	0.5037	0.5121	0.6021	-0.0084***	-0.0984***

Notes: 1. Column A lists the sample structure for the compulsory liability insurance sample.

2. Column B lists the sample structure for the extended liability insurance sample.

3. Column C lists the sample structure for the subgroup with mileage information in the compulsory liability insurance sample.

4. *** represents significance at the 1% level, ** represents significance at the 5% level, and * represents significance at the 10% level.

Table 3 Claim structure and mileage usage of green vehicles and non-green vehicles

	Compulsory liability insurance						Extended liability insurance					
	Whole Group			*Subgroup*			*Whole Group*			*Subgroup*		
	green	*Nongreen*	*Nongreen (equivalent)*	*green*	*Nongreen*	*Nongreen (equivalent)*	*green*	*Nongreen*	*Nongreen (equivalent)*	*green*	*Nongreen*	*Nongreen (equivalent)*
claim times:												
Year 2011												
=0	0.9643	0.9818	0.9834	0.9844	0.9852	0.9841	0.6229	0.8280	0.8750	0.5556	0.6354	0.6253
=1	0.0357	0.0182	0.0166	0.0156	0.0149	0.0159	0.3158	0.1547	0.1018	0.3333	0.3021	0.3089
≥2							0.0613	0.0173	0.0232	0.1111	0.0625	0.0658
Year 2012												
=0	0.9682	0.9861	0.9850	0.9828	0.9847	0.9859	0.6050	0.8388	0.7911	0.4946	0.5671	0.5496
=1	0.0318	0.0139	0.0150	0.0172	0.0153	0.0141	0.3407	0.1458	0.1878	0.4436	0.3792	0.3896
≥2							0.0543	0.0154	0.0211	0.0618	0.0537	0.0608
mileage usage:												
Year 2011												
mean				49.73	34.28	34.14				49.73	34.28	34.15
90pct				72.99	61.95	60.87				72.99	61.95	60.88
75 pct				58.01	43.52	43.55				58.01	43.52	43.54
50 pct				44.30	30.07	30.32				44.30	30.07	30.32
25 pct				23.61	18.06	18.30				23.61	18.06	18.30
10 pct				13.00	10.60	10.84				13.00	10.58	10.84
Year 2011												
mean				51.21	39.73	40.14				51.29	39.71	40.13
90pct				92.22	71.29	71.67				92.22	71.24	71.57
75 pct				62.52	48.97	49.66				62.52	48.94	49.65
50 pct				41.13	30.73	32.07				41.20	30.72	32.07
25 pct				24.66	17.50	18.31				24.66	17.50	18.37
10 pct				13.00	9.45	9.96				13.27	9.44	9.94

Note: The mileage information is expressed in kilometers driven per day.

Table 4 The claim probability and claim amount in compulsory liability insurance contracts

	Instrumental variable Probit model				Treatment effect model			
	Compulsory		equivalent		Compulsory		equivalent	
	gveh	*clm*	*gveh*	*clm*	*gveh*	*clmamt*	*gveh*	*clmamt*
constant	-0.0283***	0.2208***	-0.0803***	0.2839***	-3.0425***	111.5455	-2.3847	219.7885**
green		1.2887		0.6913		23.4136		61.9296
income	4.39E-07***		1.26E-06***		2.23E-5***		2.49E-5***	
edu	0.0006**		0.0023***		0.1760**		0.1411**	
female	-0.0005	-0.0084	-0.0005	-0.0422*	-0.0441	-8.3805*	-0.0469	-17.3102*
married	-0.0054***	0.0193	-0.0055***	0.0178	-0.5994***	0.2106	-0.3943***	-1.0724
age2530	0.0017	-0.1652***	0.0096	-0.2189***	-0.3635	-31.6959*	-0.3669	-109.2094**
age3060	0.0053***	-0.1900***	0.0168**	-0.1543***	0.3905	-25.9251*	0.1233	-91.5800**
ageabv60	0.0054***	-0.1448***	0.0133***	-0.1855***	0.3096	-23.6006	-0.0546	-95.7330**
city	0.0021***	0.0070*	0.0084***	0.0116**	-0.1051	-9.0599**	-0.1897	-9.7209**
north	-0.0007	0.0831***	-0.0009	0.1266**	0.0400	-31.4989***	0.1114	-31.4696***
south	-0.0012	0.0391	-0.0047	0.0016	0.2221	-23.3679**	0.3074	-38.9317**
central	0.0006	0.0602	0.0001	0.0309	0.3449	6.7302	0.4498	9.9263
carage0	0.0352***	-0.0305	0.1119***	-0.0934	0.0109***	9.2074	0.0438***	10.0112
carage1		0.0526		0.0979		7.9434		-14.3551
carage2		0.0271		0.0480		15.8618**		11.1804
carage3		0.0258		0.0538		7.8468		-17.5328
carage4		0.0193		0.0376		9.3428		-0.5602
bm	-0.0159***	0.4453***	-0.0457***	0.5566***	-0.0156***	-1.3381	-0.0367***	-1.9458
veh_m	-0.0046***	0.0111	-0.0224***	0.0132	-0.0644**	-5.2284	-0.0047**	-14.8877
veh_l	0.0247***	-0.1651***	0.0766***	-0.1443***	0.0362***	1.4698	0.0636***	1.6487
sedan		-0.1410***		-0.5586***		-7.8840		-19.4706
IMR						-11.3628***		-36.4440***
Haus. test	0.0018		0.0088			0.0031		0.0003
J test	0.4130		0.8449					
AR test	0.1903		0.8592					

Notes: 1. In the above table, we report the p-values for the Hausman test, J test, and AR test.

2. In the regressions of the whole sample, we have also controlled the information regarding the vehicle brands *tramak_i=n, f, h, t, mz, mts*. However, we do not report this information in the table to maintain confidentiality.

Table 5 The claim frequency and claim amount in extended liability insurance contracts

	Instrumental variable NB model				Treatment effect model			
	Extended		equivalent		Extended		equivalent	
	gveh	*clmnumber*	*gveh*	*clmnumber*	*gveh*	*clmamt*	*gveh*	*clmamt*
constant	-3.2352	3.0282***		7.6946**	-2.0771***	99.1370***	3.9813***	72.8015***
green		0.1233**		0.0897***		54.7052*		30.3832
avincm	3.99E-05***		0.0001***		1.46E-05***		1.72E-05***	
edu	0.0306***		0.0025**		0.0102***			
channel_D								
city	0.5903***	0.0452***	0.6208***	0.0508**	0.2099***	-5.3296***	0.1207***	-4.5637***
female	-1.2141***	0.1179***	-1.7483***	0.1912***	-0.1390***	-6.4020***	-0.1454***	-2.5871
married	-2.1230***	0.0744***	-0.6944***	-0.0831***	-0.4672***	2.0546	0.0680	-2.1395
age2530	-1.6600	-0.1055	-1.8524	0.0850	0.1739	-34.3205***	0.0284	-14.6970***
age3060	1.9203**	-0.3358	3.9143**	0.1678	0.3990*	-37.5377***	0.0539**	-14.6402***
ageabv60	1.6184*	-0.3373	3.9025**	0.1967	0.4786*	-31.9305***	0.0306*	-14.2620***
north	-0.2735	0.0956	0.0766	0.1376***	0.0277**	-12.3055***	0.0562	-3.5033
south	-1.2770**	-0.0580	-0.6178**	-0.0095	-0.0094	-2.9370	-0.0071	0.9151
central	-1.0146**	-0.0873	-0.7080**	-0.0098	0.0653	-1.9006	0.0394	4.0898
carage0	0.0373***	1.4707***	0.0207**	1.6054***	0.0886***	-10.6022***	0.0276**	-11.5498**
carage1		0.9358***		1.0523***		-5.2958**		-7.1597
carage2		0.5607***		0.7203***		0.0128		0.2079
carage3		0.2949***		0.4105***		0.0375		-3.0683
carage4		0.1551***		0.2464***		3.9374		1.9436
bm	-0.0517***	0.6553***	-0.0458***	0.5956***	-0.0859***	1.5420	-0.0625***	1.5644
veh_m	0.0021	0.0475	0.0033	-0.0085	0.0098	1.6837	0.0071	-1.2161
veh_l	0.0249***	0.1809***	0.1627***	0.1666***	0.0379***	15.5759***	0.0329***	20.3631***
sedan		0.3184		-0.4295		15.2627***		22.6282***
IMR_{extend}	-6.6730***	-0.8686***	-14.1772***	-0.1298***	-0.4937***	3.4994***	-5.7527***	71.7447**
IMR						-18.5108***		-11.2942***
Haus. test	<0.0001		<0.0001			<0.0001		<0.0001
J test	0.4110		0.4178					
AR test	0.5912		0.6193					

Notes: 1. In the above table, we report the p-values for the Hausman test, J test, and AR test.

2. In the regressions of the whole sample, we have also controlled the information regarding the vehicle brands *tramak_i*=*n*, *f*, *h*, *t*, *mz*, *mts*. However, we do not report this information in the table to maintain confidentiality.

Table 6 The mileage usage analysis in the compulsory liability insurance sample

	Compulsory		**Equivalent**	
	gveh	*mileage*	*gveh*	*mileage*
constant	-2.2979***	-282.0979*	-1.9313***	52.4555
green		18.4338***		26.4312***
income	8.63E-07***		0.0000***	
edu	0.9326**		2.8684**	
access				
female	-0.2304***	9.9970	-0.1386***	-1.2493
married	-0.1917**	-76.3978**	-0.2931***	-3.7578*
age2530	0.2417	-10.6975	0.2139	-29.2120**
age3060	0.4746*	8.1174	0.4172*	-30.2625**
ageabv60	0.5679*	4.1331	0.5020*	-34.9713***
city	0.1177***	-14.8800**	0.0977**	-1.6729*
north	-0.1639	-18.5399**	-0.2866**	-6.2994**
south	-0.2915**	-3.6125	-0.3966***	-3.1534
central	-0.2434**	-1.7729	-0.3555***	-8.0645
carage0		9.7815**		45.8201***
bm		-7.2862		-8.0520
veh_m		-2.2279		-0.8725
veh_l		-11.5470		-4.2188
sedan		-2.6874		-26.5858
IMR_{repair}	0.0236*	111.5178**	0.4400***	24.3695***
IMR		34.3349**		61.1767***

Table 7 The mileage usage analysis in the extended liability insurance sample

	Extended		Equivalent	
	green	*mileage*	*green*	*Mileage*
constant	-1.6346**	64.4027	-1.2134	20.8830
green		19.3989***		23.8371***
income	1.19E-05***		8.02E-06**	
edu	0.8936**		2.2944***	
access				
female	-0.3422***	-2.6993	-0.2998**	-5.6129
married	-0.2181**	0.2672	-0.3269***	-6.8771
age2530	0.0189	-14.9173*	-0.0919	-35.4131**
age3060	0.1028	-15.9657	0.0023	-36.0460*
ageabv60	0.2184	-21.5591	0.1054	-40.0582*
city	0.1326**	-0.6241***	0.1336**	-0.1323
north	-0.1803	1.5971	-0.2722**	-6.7807
south	-0.3719***	-0.4552	-0.4887***	-7.4832
central	-0.3362	-1.7147	-0.4493***	-12.5305
carage0		6.9932***		43.6855
bm		-6.2278		-0.3389
veh_m		7.2285***		1.6138
veh_l		-0.2318		-6.5967
sedan		-1.3989		13.4589
IMR_{repair}	0.0074*	3.3776***	0.4904***	3.6135***
IMR_{extend}	-0.7596***	1.1309**	-0.9311***	1.6331**
IMR		34.8524**		70.3251***

Table 8 The claim probability and claim amount, based on mileage, in the compulsory liability insurance contracts

	Instrumental variable Tobit model				Treatment effect model			
	Compulsory		equivalent		Compulsory		equivalent	
	gveh	*clm_pkkm*	*gveh*	*clm_pkkm*	*gveh*	*clmamt_pkkm*	*gveh*	*clmamt_pkkm*
constant	0.5531**	0.4371**	0.2479*	0.3727**	4.9753*	10.5794***	6.6828**	96.9934***
green		-0.0574***		-0.0105**		-1.4165***		-1.2580***
income	1.04E-06***		4.12E-07***		0.0029***		0.0039***	
edu	0.0298**		0.3435***		0.5595***		0.2416***	
female	-0.0324	-0.0030**	-0.0715	-0.0103***	-0.7246***	-5.6886*	-0.8251***	-3.8752
married	0.1225	0.0151	0.1506	0.0231	0.1385**	0.7778	0.1718	8.6255
age2530	-0.0084	-0.0090*	0.0059	-0.0025*	-0.6072	-3.0638	-0.2992	-15.7637
age3060	0.0548**	-0.0091***	0.0697**	-0.0031***	0.1417**	1.9641	0.3815***	-13.4985
ageabv60	0.0535**	-0.0095*	0.0172*	-0.0105*	0.1358*	-2.6434	0.3051***	-13.7772
city	0.0296*	0.0039***	0.0849**	0.0150***	0.9185**	-1.1190**	0.3773***	-1.2226**
north	0.0296*	0.0039***	0.1707*	0.0297***	0.2233	-1.1955**	0.1928	-2.8469***
south	-0.0095	-0.0001	-0.0468	-0.0057	-0.2625	-3.1049***	-0.2483	-8.5207***
central	-0.0119	-0.0020	-0.0493	-0.0069	-0.6019	2.6229	-0.6800	1.6073
carage0	0.2695	-0.0367	-1.6786	-0.2746	0.3781*	5.2159	0.4400*	-48.1777
bm	-0.0697**	0.0024***	-0.1857***	0.0233***	-0.2095**	4.8879	-0.2745*	-23.6273
veh_m	-0.0112	0.0017	-0.0299	0.0055	0.1526	-1.0074	0.1714	-5.6274
veh_l	0.2419**	-0.0231**	0.4931***	-0.0819***	0.8512***	-3.5083	0.9339***	-2.3533
IMR_{repair}	-0.1928**	-0.0235***	-0.8571**	-0.1421***	-0.2230***	-1.9415***	-0.3782***	-1.7189***
IMR						-19.2200***		-59.1060***
Haus. test	0.0006		0.0032			<0.0001		<0.0001
J test	0.6604		0.3236					
AR test	0.4649		0.1884					

Note: 1. In the above table, we report the p-values for the Hausman test, J test, and AR test.

Table 9 The claim probability and claim amount, based on mileage, in the extended liability insurance contracts

	Instrumental variable Tobit model				Treatment effect model			
	Extended		equivalent		Extended		equivalent	
	gveh	*clm_pkkm*	*gveh*	*clm_pkkm*	*gveh*	*clmamt_pkkm*	*gveh*	*clmamt_pkkm*
constant	0.1456*	-2.7430***	0.0452**	-1.7510**	0.6645	2.6320**	0.8245	7.0541***
green		-0.5109***		-0.3278***		-1.7005***		-1.1508***
income	7.86E-07***		1.86E-07***		6.18E-06***		6.00E-07***	
edu	0.0791***		0.3457***		0.2985***		0.5547***	
female	-0.0509***	-0.0218**	-0.0099*	-0.0703***	-0.0910**	0.4538	-0.0310*	0.9481
married	0.1936*	0.3536	-0.0160	-0.0937	-0.0024	-0.9105	-0.0150	-0.2781
age2530	-0.0383	0.0298	-0.0432	0.1043	-0.0449	0.2881	0.2930	0.5213
age3060	0.1199**	-0.1128***	0.0516***	-0.0423**	0.47871**	0.4198	0.2853**	0.8673
ageabv60	0.1194**	-0.1351***	0.0374**	-0.0130*	0.4949**	0.9290	0.3046***	0.8196
city	0.0409**	0.0938***	0.0118***	0.0440***	0.2702***	-0.2495**	0.2108***	-0.2438**
north	0.0311	0.1291***	0.0137	0.1116***	0.0454	-0.7102***	-0.0009	-0.6138***
south	-0.0080	0.0670	-0.0403**	0.1150	-0.0252*	-0.5042**	-0.0287**	-0.6972**
central	-0.0209	-0.0283	-0.0299*	-0.0117	-0.0284*	0.0856	-0.0224*	0.2486
carage0	0.4432**	-0.9464	0.0567***	-0.7638	0.9916***	2.3190	0.3478***	5.8857
bm	-0.1024***	0.0145***	-0.0295***	0.1552***	-0.3857***	0.5440	-0.0471*	0.1809
veh_m	0.0154	0.0273	0.0073	-0.1283	0.0334	-0.3330	0.3072	-0.0582
veh_l	0.2612***	-0.1253***	0.3806***	-0.6047***	0.1726***	-0.1875	0.2538	-0.2423
IMR_{repair}	-0.2778***	-0.5195***	-0.0185*	-0.4878***	2.7511***	-2.9255*	2.4966***	-3.8310*
IMR_{extend}	-0.0647*	-0.3075**	-0.0234**	-0.4925***	-1.0439***	-0.3503**	-1.0241***	-0.4095***
IMR						0.2438***		0.3789***
Haus. test	0.0011		0.0005			<0.0001		<0.0001
J test	0.3057		0.4020					
AR test	0.6080		0.5707					

Note: 1. In the above table, we report the p-values for the Hausman test, J test, and AR test.

Appendix 1

Table A1

	green	*non-green*		*difference*
		whole	equivalent	
	(A)		(B)	(A)-(B)
female	0.6466	0.6176	0.6219	0.0247***
married	0.6486	0.6816	0.6348	0.0138
age2025	0.0042	0.0131	0.0077	-0.0035*
age2530	0.0319	0.0542	0.0352	-0.0033
age3060	0.8744	0.8443	0.8493	0.0251***
ageabv60	0.0892	0.0880	0.1077	-0.0185***
city	0.6972	0.6195	0.6283	0.0689***
north	0.5433	0.4955	0.4885	0.0548***
south	0.2431	0.2709	0.2795	-0.0364***
central	0.1848	0.1809	0.1872	-0.0024
carage0	0.5723	0.1492	0.1886	0.3837***
carage1	0.1815	0.1615	0.1760	0.0055***
carage2	0.1000	0.1449	0.1411	-0.0411***
carage3	0.0714	0.1406	0.1351	-0.0637***
carage4	0.0253	0.1312	0.1205	-0.0952***
veh_m	0.0004	0.2579	0.1984	-0.1980***
veh_l	0.5692	0.2308	0.2537	0.3155***
sedan	0.9949	0.8799	0.9431	0.0518***
tramak_n	0.0000	0.0107	0.0000	0.0000
tramak_f	0.0000	0.0697	0.0000	0.0000
tramak_h	0.0011	0.0893	0.0024	-0.0013*
tramak_t	0.9953	0.4590	0.9874	0.0079***
tramak_mts	0.0000	0.0526	0.0000	0.0000
tramak_mz	0.0000	0.0475	0.0000	0.0000
avincm	29.3388	27.2617	27.4126	1.9262***
edu	0.0792	0.0683	0.0699	0.0093***
cont	0.2788	0.5230	0.5236	-0.2448***

Notes: (1) The green vehicles account for 1.51% of the whole sample, and 6.66% of the equivalent sample.

(2) The significance of the difference between (A) and (B), and between (A) and (C) is tested by the t test. * represents significance at the 10% level, ** represents significance at the 5% level, and *** represents significance at the 10% level.

Appendix 2

Table A2

	Compulsory		Equivalent	
	$select_{extend}$	$select_{repair}$	$select_{extend}$	$select_{repair}$
intercept	-1.1075***	-2.7528***	-1.0505***	-2.0938***
income	-2.63E-07***	9.71E-07***	-1.88E-07***	6.81E-07***
edu	0.0589***	0.2804**	0.0326**	2.1366*
channel_D	0.8032***		0.9353***	
access		0.0006*		0.0318***
female	0.3853***	0.1594***	0.4632***	0.1623***
married	0.1785	-0.9140***	0.1098	-0.4345***
age2530	0.5099	0.0480	0.5964	0.2570**
age3060	0.9694**	0.2996***	0.9860**	0.2178*
ageabv60	0.8890**	0.3223***	0.9328**	0.1927
city	0.0442***	-0.1801***	0.0682***	-0.0929***
north	0.0748	-0.2508***	0.0700	-0.4487***
south	0.2918*	-0.0405	0.2147*	-0.0476
centrol	0.3446*	-0.0004	0.2921*	-0.2491***
carage0		2.2245***		2.7962***
veh_m		0.1161***		0.0722**
veh_l		0.1417***E		0.3321***