Mileage-Based Pricing in Vehicle Insurance Products: China as a case study

CHEN Dongmei *
Duan Baige **

Abstract: Since there is no mileage statistics in the database of almost all property insurance companies in China currently, the paper proposes to study the impact of driving mileages on net premiums of vehicle insurance innovatively. The paper makes use of the data of 31 provinces, municipalities and autonomous regions in China from 2000 to 2009 and makes an empirical analysis to study the long-term stable equilibrium relationship between the accident losses and highway mileages by using the panel unit root test and co-integration test of econometric methods. The result shows that there is a significant positive correlation relationship between the accident losses and highway mileages. To some extent, the stable relationship can be regarded as a valid reference of the relationship between net premiums of vehicle insurance and driving mileages. Moreover, the paper proposes some feasible programs to consider the relationship between net premiums of vehicle insurance and driving mileages, and provides further proposal for mileage-based pricing in vehicle insurance products of China. The results of the paper have important theoretical significance and practical value for driving mileages to be introduced and developed into GLM pricing in China's vehicle insurance, and also provide a reference to settle our transportation, energy and environmental problems.

Keywords: generalized linear models; driving mileages; vehicle insurance; cross-subsidies; unit root test; co-integration test; positive externalities

* Dongmei Chen, Associate Professor and Director, Department of Insurance, Fudan University School of Economics, 600 Guoquan Road, Shanghai, China, 200433; Email:dmchen05@gmail.com; Phone: 0086-21-65643750, Fax:0086-21-65647719

** Baige Duan is a PhD student of Department of Risk Management and Insurance at the School of Economics, Nankai University majoring in Actuarial Science, Email: asabaigefsa@gmail.com

1. Introduction and Motivation

In the field of non-life actuarial science, insurance data often don't follow normal distribution, generalize linear models (GLMs) are very suitable for the analysis of such data. In the pricing of non-life insurance products, some estimation methods which were very early and have been widely applied are really the applications of GLMs, such as Bailey-Simon method, or minimum deviation method; the marginal summation method; least squares estimation method; and some other intuitive approaches; etc. The relevant literatures can refer to Brown(1988), Schmidt & Wünsche(1998), Mildenhall(1999), Feldblum & Brosius(2002), Fu & Wu(2007), etc. Recently, some relevant materials and monographs which based on the application of GLMs in the international non-life actuarial science have also been published in succession. de Jong & Heller(2008) first described how to apply GLMs for insurance data, which provided a lot of numerical illustrations and had a detailed analysis. The latest monograph can refer to Ohlsson & Johansson (2010), which described the various extensions of GLMs in detail, such as GAMs, and can be regarded as a useful supplement of de Jong & Heller(2008).

In the 1990s, GLMs were introduced into non-life insurance pricing by British actuaries. The following two decades, GLMs have achieved considerable development in the non-life pricing practices in many countries, and have been become the standard method currently. With the further opening up of China's insurance market, domestic insurance companies are bound to learn overseas advanced actuarial techniques. With the active and rapid development of China's property insurance market, China Insurance Regulatory Commission (CIRC) firsts began to consider the vehicle insurance that the main non-life insurance business, released the 2010(619) notice of property and casualty insurance, i.e. carry out the pilot reform of commercial vehicle insurance pricing mechanism in Shenzhen. Therefore, the market-oriented vehicle insurance rates once again become the focus of insurance industry. Facing the new challenges of vehicle insurance pricing and upgrade of underwriting capacity, property and casualty insurance companies should focus on how to efficiently analyze the impact factors of rates of vehicle insurance data, such as vehicle's factors, people's factors and regional factors, and analyze carefully the impact of various risk factors on loss frequency and loss severity significantly, so as to enhance pricing capabilities of vehicle insurance and identify profitable business by automated underwriting of GLMs, further to optimize the marketing strategy and flexibility to improve business structure. These studies are effective ways of property and casualty insurance companies to maintain the performance of solid growth in the fierce competition. However, through the discussion and exchange of the authors and some actuaries of domestic property and casualty insurance companies, it is generally agreed that driving mileage is positively correlated with the risk exposure of vehicle. The mileage factor should be considered in vehicle insurance pricing, but there is no mileage statistics in the database of almost all domestic property insurance companies currently. Faced with the situation, it has become an urgent problem of how to incorporate the factor into vehicle insurance pricing.

In addition, the traditional pricing model of vehicle insurance does not fully reflect the fairness of premiums and there have been the phenomenon of "high premium low compensation" in China. The reason is as follows. In the case of assuming other conditions remain unchanged, the risks of the low-mileage drivers are generally lower than the high-mileage drivers, yet their premiums charged are the same, which leads to a phenomenon that the low-mileage drivers provide cross-subsidies for the high-mileage drivers. Moreover, from an economics perspective, the model may also be prompted to increase the risks of the drivers by increasing mileages. That is, in the case of same premiums, the drivers would tend to increase mileages so as to make their own utility maximization. However, Mileage-based pricing of vehicle insurance product can not only reasonably reflect the insured's insurance cost so as to show the fairness of premium charged, but also alleviate the traffic congestion effectively, reduce energy consumption, reduce environmental pollution and greenhouse gas emissions, and resulting in other comprehensive social benefits to some extent. Therefore mileage-based insurance has become a popular product in vehicle insurance market of developed countries in Europe and America.

In view of these motivations, the paper proposes to study the impact of driving mileages on net premiums of vehicle insurance innovatively. The paper makes use of the macro-statistics data of highway mileages, traffic accidents and the accident losses based on China Statistical Yearbook from 2001 to 2010, establishes the econometric model using the data of 31 provinces, municipalities and autonomous regions in China from 2000 to 2009 and makes an empirical analysis to study the long-term stable equilibrium relationship between the accident losses and highway mileages by using the panel unit root test and co-integration test of econometric methods. The result shows that there is a significant positive correlation relationship between the accident losses and highway mileages. To some extent, the stable relationship can be regarded as a valid reference of the relationship between net premiums of vehicle insurance and driving mileages. In vehicle insurance pricing based on GLMs, property and casualty insurance companies can make use of the available and specific vehicle's, people's and regional information, for example, they can select vehicle types, age of drivers, the respective district of vehicles as ratemaking variables, analyze the impact of these variables on loss frequency and loss severity. On this basis, they can determine the rates, develop vehicle insurance products with competitive advantage. When the expiration of insurance period, they can also make use of the difference between displayed mileages of vehicle odometer in the period of insurance and estimated average mileages[1] of econometric methods to adjust net premiums of vehicle insurance within a reasonable range of elasticity. Moreover, they can minute the actual driving mileages to the database of property and casualty insurance companies according to the insured's odometer information within the insurance period. Once the data are mature, driving mileages will be added to ratemaking variables and pricing based on GLMs. These studies have important theoretical support and practical reference for domestic property and casualty insurance companies to further explore the independent pricing based on GLMs in China's vehicle insurance, and provide an important theoretical basis for the insurance industry to develop new pricing software of property and casualty insurance.

2. Study of the Impacts of Highway Mileages on Accident Losses

2.1 Data Source and explanation

Modeling data of the paper derived from China Statistical Yearbook, the samples are annual data of 31 provinces, municipalities and autonomous regions in China from 2000 to 2009, including highway mileages, traffic

[1] The average mileage is calculated indirectly through the relationship between net premiums of vehicle insurance and driving mileage in the established econometric model.

accidents and the accident losses. In order to more clearly describe these data and further determine the specific form of the model reasonably, figure 1 and figure 2 show the corresponding three-dimensional maps of the accident losses (unit: million), highway mileages (unit: kilometer) and traffic accidents (unit: number) respectively. Among them, the left in figure 1 draws the accident losses of all regions at current prices according to China Statistical Yearbook, considering 2000 as the base period, the right in figure 1 draws the accident losses of all regions in 2000, which adjusted current prices to base period's prices drawn the retail price index by China Statistical Yearbook. In addition, two figures denote provinces using these digital 1-31 in order to facilitate the description. According to the order of China Statistical Yearbook, these digital denote Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shanxi, Gansu, Qinghai, Ningxia, Xinjiang, respectively.

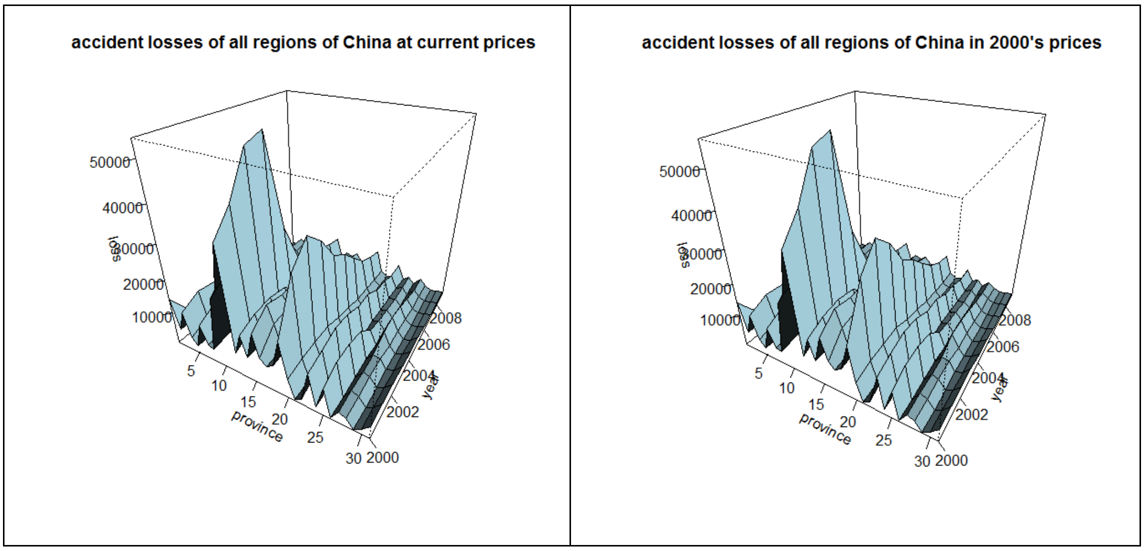

Fig. 1 The Accident Losses in the 31 Regions of China (2000-2009)

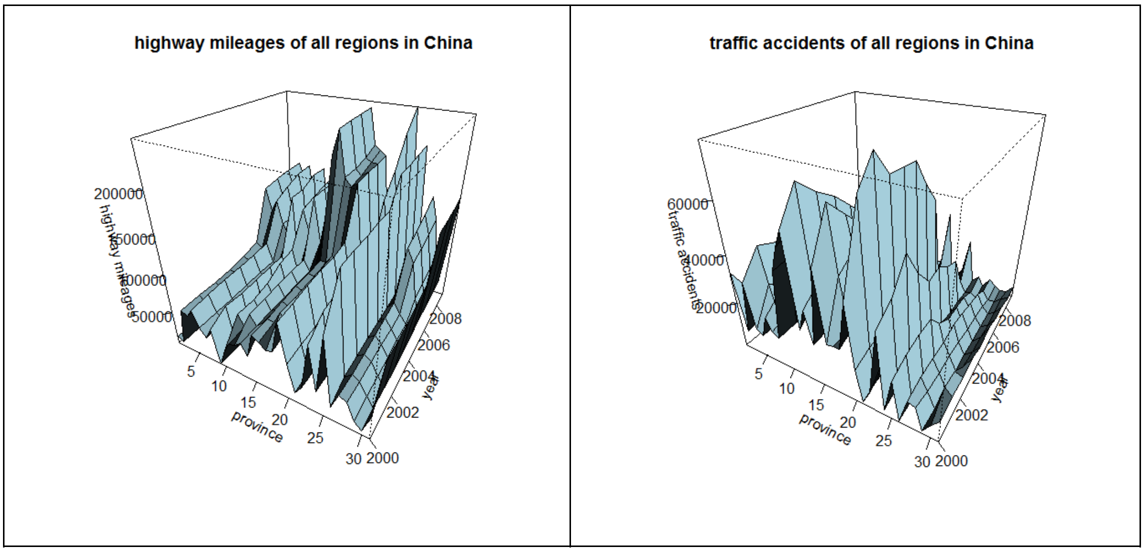

Fig. 2 Highway Mileages and Traffic Accidents in the 31 Regions of China (2000-2009)

2.2 Variable Selection and Model Description

2.2.1 The selection of dependent variable and explanatory variables

Given that the original intention of the paper is to study the impact of driving mileages on net premiums of vehicle insurance, so we select the accident losses after adjustment as dependent variable, highway mileages and traffic accidents are explanatory variables. These variables are expressed as $Loss$, $Mileage$, $Accidents$. Also, the actual macroeconomic variables sequences generally show such feature, i.e. with the changes in the values of explanatory variables, the difference in the value of dependent variable will be more generally. In order to eliminate the impact of increment heteroscedasticity that may arise, the paper uses the natural logarithms[2] of annual $Loss$, $Mileage$, $Accidents$ data of 31 regions, and further obtains the regional panel data sequences, including $\ln Loss$ 、 $\ln Mileage$ 、 $\ln Accidents$.

2.2.2 The selection of specific explanatory variables

In the actual modeling process, dependent variable not only is affected by the quantitative variables, but also may be subject to the qualitative factors. For example, it may need to consider the impact of gender, different historical periods, seasonal differences, regional differences and other factors according to the specific problem. Figure 1 shows that the accident losses of 31 regions are in the significant growth until 2003 and apparent downward trend after 2003. Figure 2 shows that highway mileages have been a significant increase after 2006, traffic accidents of almost all regions had been showed an upward trend before 2003 and declined slightly from 2003 to 2005, and a downward trend generally after 2006. To demonstrate the overall characteristics, figure 3 shows the trends of annual the accident losses after adjustment, highway mileages and traffic accidents from 2000 to 2009 in China. Among them, the horizontal axis represents year, the primary vertical axis (on the left vertical axis) used to describe annual highway mileages, and the secondary vertical axis (on the right vertical axis) used to describe annual accident losses and traffic accidents[3]. From a macro perspective, figure 3 shows that these features are closely related to national economy, safety conditions of road and traffic, sound legal system of China. In addition, these features also are consistent with the development of China's vehicle insurance business. Before 2003, vehicle insurance rates are unified pricing based on several vehicle's factors; from January 2003 to July 2006, vehicle insurance rates for market-oriented reforms were implemented, a priori classification rate system was established based on some factors of people and vehicle; since July 2006, with the implementation of traffic accident compulsory liability insurance, rates were again returned to the unified model, and the classification rate system was further revised with a combination of vehicle, of people, of region and other factors.

In order to reflect these features, the paper adds a dummy variable (D) to measure the change in the intercept of the model, adds the product of quantitative variables and a dummy variable to measure the change in the slope of the model. ($\ln Mileage \times D$, $\ln Accidents \times D$). And the dummy variable is defined as follows:

[2] Generally, in addition to solve the problem of heteroscedasticity, there are the following two aspects of the role of logarithmic variables mainly, the first is the nonlinear problems involving exponential functions and so on can be transformed into the linear problems that easy to deal with, and second is that if the logarithmic sequence is still non-stationary, then the result approaches the growth rate after difference and has certain economic implication.

[3] The longitudinal coordinate scale of the secondary vertical axis is implemented of the conversion. i.e., located in inside coordinate scale used to describe the accident losses, located in outside coordinate scale used to describe traffic accidents. Therefore, we can draw accurately the trends of three variables over time in the same figure.

$$D = \begin{cases} 0, & 2000 \le t < 2006 \\ 1, & 2006 \le t \le 2009 \end{cases} \quad (1)$$

It is worth mentioned, the paper uses a single dummy variable to two explanatory variables both highway mileages and traffic accidents in measuring the changes in the intercept and slope of the model. The reason is as follows. The paper also applies the two dummy variables (D_1, D_2) and two product terms ($\ln Mileage \times D_2$, $\ln Accidents \times D_1$) showed by the following second and third formulas in the empirical analysis, and conducts the co-integration test and estimated parameters of the models. We choose the single dummy variable ultimately by further balancing the results of estimated parameters and important statistical indicators under the two approaches.

$$D_1 = \begin{cases} 0, & 2000 \le t < 2003 \\ 1, & 2003 \le t \le 2009 \end{cases} \quad (2)$$

$$D_2 = \begin{cases} 0, & 2000 \le t < 2006 \\ 1, & 2006 \le t \le 2009 \end{cases} \quad (3)$$

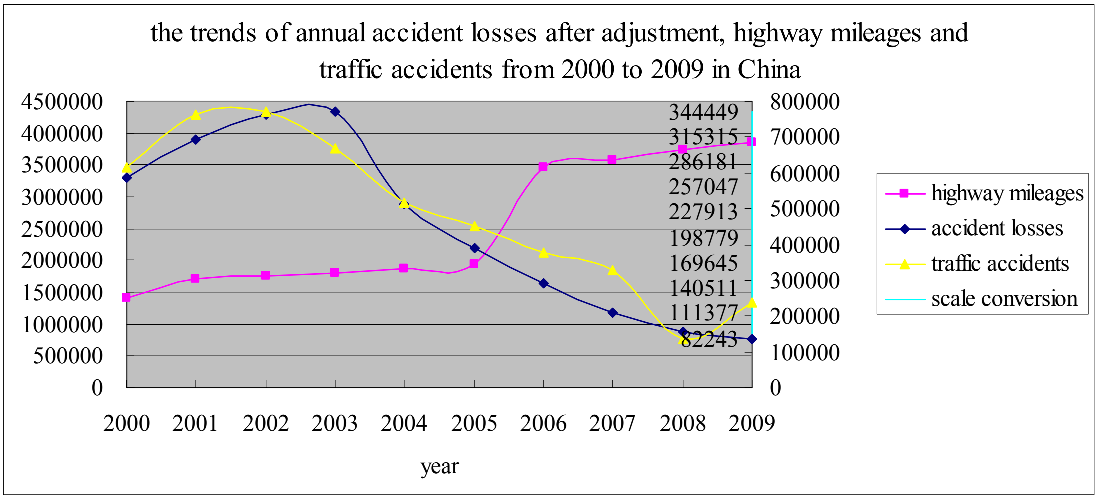

Fig. 3 Annual Accidents Losses, Highway Mileages and Traffic Accidents of China (2000-2009)

2.2.3 The basic form of the model

The panel data not only combine the characteristics of time series data and cross-sectional data, so as to increase the observations significantly and further improve the sampling accuracy of estimator, but also obtain more dynamic information than a single cross-sectional data by modeling. Moreover, we can get consistent parameter estimates and even efficient estimators in the fixed effects model of panel data. Therefore, the paper proposes to select the panel data for modeling. The basic forms of the optional models are as follows.

1. Pooled Model

According to the paper, the pooled model can be expressed as:

$$\begin{aligned} \ln Loss_{it} &= \alpha + \beta_1 \ln Mileage_{it} + \beta_2 \ln Accidents_{it} + \beta_3 D \\ &+ \beta_4 \ln Mileage_{it} \times D + \beta_5 \ln Accidents_{it} \times D + \beta_6 \ln Loss_{i,t-1} + \varepsilon_{it} \end{aligned} \quad (4)$$

2. Fixed Effects Model

The fixed effects model can be divided into three types, i.e. entity fixed effects model, time fixed effects model, time and entity fixed effects model. According to the paper, entity fixed effects model can be expressed as:

$$\begin{aligned} \ln Loss_{it} &= \alpha_i + \beta_1 \ln Mileage_{it} + \beta_2 \ln Accidents_{it} + \beta_3 D \\ &+ \beta_4 \ln Mileage_{it} \times D + \beta_5 \ln Accidents_{it} \times D + \beta_6 \ln Loss_{i,t-1} + \varepsilon_{it} \end{aligned} \quad (5)$$

According to the paper, time fixed effects model can be expressed as:

$$\ln Loss_{it} = \gamma_t + \beta_1 \ln Mileage_{it} + \beta_2 \ln Accidents_{it} + \beta_3 D + \beta_4 \ln Mileage_{it} \times D + \beta_5 \ln Accidents_{it} \times D + \varepsilon_{it} \quad (6)$$

According to the paper, time and entity fixed effects model can be expressed as:

$$\ln Loss_{it} = \alpha_0 + \alpha_i + \gamma_t + \beta_1 \ln Mileage_{it} + \beta_2 \ln Accidents_{it} + \beta_3 D + \beta_4 \ln Mileage_{it} \times D + \beta_5 \ln Accidents_{it} \times D + \varepsilon_{it} \quad (7)$$

Therein, the first-order lagged terms of dependent variable ($\ln Loss_{i,t-1}$) is added to the model in the fourth and fifth formulas. Its purpose is to overcome the autocorrelation of dependent variable sequence. However, because of the autoregressive terms (AR) aren't emerge in the time fixed effects model, first-order lagged terms of dependent variable can't be added to the model in the sixth and seventh formulas.

3. Random Effects Model

The random effects model can be divided into three types, i.e. entity random effects model, time random effects model, time and entity random effects model. The basic forms of the corresponding models are the same as the fifth, sixth, seventh formulas. The only difference is that the distribution of intercept term (i.e. constant term) is not relevant to explanatory variables in random effects model, but the distribution of intercept term (i.e. constant term) is relevant to explanatory variables in fixed effects model.

2.3 Model Testing, Parameter Estimation and Analysis of Results

2.3.1 Unit Root Test of Panel Data

As a result of the regression model using non-stationary economic variables can bring the problem of spurious regression, we first apply unit root test of panel data before modeling, so as to ensure each variable in the model is stable. There are many methods to conduct unit root test of panel data. According to the different scope of the application, these methods can be divided into two types. The first one is suitable for entity of panel data with the same root, and the second one is suitable for entity of panel data with the different root. Therein, the null hypotheses of the two types are with unit root, alternative hypotheses are without unit root. The paper selects LLC testing, IPS testing, In Choi testing (also known as Fish-ADF testing), Fisher-PP testing to apply unit root test on variables. In the four test methods, LLC testing is suitable for the same root, the other three methods are suitable for the different root. The test results of the paper are shown in table 1.

Table 1 Unit Root Test of Panel Data

test methods / variables	LLC testing	IPS testing	Fisher-ADF testing	Fisher-PP testing
$Loss$	-6.35（0.00）***	1.88（0.97）	51.81（0.82）	31.43（1.00）
$\ln Loss$	1.20（0.88）	5.17（1.00）	23.05（1.00）	13.23（1.00）
$\Delta \ln Loss$	-11.93（0.00）***	-4.92（0.00）***	136.21（0.00）***	186.98（0.00）***
$Mileage$	1.03（0.85）	4.95（1.00）	18.44（1.00）	22.94（1.00）
$\ln Mileage$	-1.35（0.09）*	3.24（1.00）	33.97（1.00）	43.11（0.97）
$\Delta \ln Mileage$	-23.40（0.00）***	-8.74（0.00）***	190.60（0.00）***	239.78（0.00）***
$Accidents$	-7.34（0.00）***	1.91（0.97）	46.78（0.92）	40.77（0.98）
$\ln Accidents$	2.71（1.00）	5.87（1.00）	19.66（1.00）	35.11（1.00）
$\Delta \ln Accidents$	-13.75（0.00）***	-8.14（0.00）***	201.20（0.00）***	266.17（0.00）***

Note: ***, **, * indicate that reject the null hypothesis at the significant level of one percent, five percents and ten percents respectively. The number on the left of parentheses indicates the value of corresponding statistics. The number in the parentheses indicates the accompanied probability of corresponding statistics. The notation (Δ) indicates the first-order difference of corresponding variable.

It can be seen from table 1, three variables and corresponding logarithm variables are not stable, i.e. there are unit root. However, the first-order differences of logarithm variables are stable, i.e. logarithm variables are first-order integration ($I(1)$). Therefore, it is consistent with the requirement of co-integration test of panel data.

2.3.2 Co-integration Test of Panel Data

The co-integration test of the model can be divided into two types by different methods in panel data. The first one is the co-integration test of panel data based on the extended two-step method of Enger-Granger (EG) in the co-integration test of time series, and the constructed test statistic after the standardization follow standard normal distribution asymptotically. The typical methods are Pedroni co-integration test method and Kao co-integration test method. Therein, the former is a class of co-integration test for heterogeneous panel data, which allows the heterogeneity of the system among different entities, the latter is a class of co-integration test for homogeneous panel data. And the second one is the co-integration test of panel data based on the extended Johansen trace statistics. The typical method is Fisher co-integration test method, which is to construct a cumulative statistic with following the chi-square distribution by the values of co-integration test of entities, in order to develop co-integration test of panel data. The null hypotheses of the two types of methods are without the co-integration relationship, alternative hypotheses are with the co-integration relationship.

The three logarithmic variables are the process of first-order integration in our model, so we can further test whether the co-integration relationship or not between dependent variable and explanatory variables. Pedroni constructs seven statistics to test the co-integration relationship of panel data, these statistics are applied to determine whether co-integration relationships or not between $\ln Loss$ and $\ln Mileage$, $\ln Loss$ and $\ln Accidents$[4]. There statistics can be divided into two types. The first one is described by within-dimension, including four statistics, i.e. $Panel\ v$ 、 $Panel\ \rho$ 、 $Panel\ PP$ 、 $Panel\ ADF$, and the second one is described by between-dimension, including three statistics, i.e. $Group\ \rho$ 、$Group\ PP$ 、$Group\ ADF$. These test results are shown in table 2.

Table 2 Test Results of Co-integration Relationship Between $\ln Loss$ and $\ln Mileage$

Test Statistic	The Value of Corresponding Statistic	Probability
$Panel\ v$ Statistic	2.58***	0.0049
$Panel\ \rho$ Statistic	-1.70**	0.0441
$Panel\ PP$ Statistic	-4.73***	0.0000
$Panel\ ADF$ Statistic	-.3.43***	0.0003
$Group\ \rho$ Statistic	1.52	0.9361
$Group\ PP$ Statistic	-.3.97***	0.0000
$Group\ ADF$ Statistic	-5.20***	0.0000

Note: ***, **, * indicate that reject the null hypothesis at the significant level of one percent, five percents and ten percents respectively.

It can be seen from table 2, six test results of the seven show that reject the null hypothesis, so we think that there is a co-integration relationship between $\ln Loss$ and $\ln Mileage$. In addition, the paper also uses Kao co-integration test method and Fisher co-integration test method, which reject the null hypothesis at the significant level of one percent respectively, i.e. there is a co-integration relationship between $\ln Loss$ and $\ln Mileage$. Therefore, we can assume that there is an equilibrium relationship of long-term stability between $\ln Loss$ and $\ln Mileage$, so as to apply regression analysis directly, can't also bring the problem of spurious regression.

2.3.3 Model Testing and Parameter Estimation

The test methods to identify the types of panel data model are mainly F test and Hausman test. Therein, F test used to test that a set of panel data should be established of a pooled model or an entity fixed effects model. The null hypothesis is established of a pooled model, alternative hypothesis is established of an entity fixed effects model. Hausman test[5] used to test that a set of panel data should be established of a random effects model or a fixed effects model. The null hypothesis is established of a random effects model, alternative hypothesis is established of a fixed effects model. Our test results are shown in table 3 and table 4.

Table 3 Results of F test

[4] Test results show that there are co-integration relationships between two explanatory variables and dependent variable. As the purpose of the paper is to study the impact of driving mileages on accident losses, we only provide the test result of co-integration relationship between $\ln Loss$ and $\ln Mileage$.

[5] Generally, Hausman test is not limited to test model and setting method of panel data. As long as involving the judgment of difference measurement of estimators in two estimation methods, we can also apply Hausman test method.

Test Statistic	Value of Statistic	Degrees of Freedom	Probability
F Statistic	20.26	（30,275）	0.0000
χ^2 Statistic	361.60	30	0.0000

表 4 Results of Hausman Test

Null Hypothesis	Value of Statistic $Chi-Sq$	Degrees of Freedom ($Chi-Sq$)	Probability
Entity Effect and Regression Variables are not relevant	48.03	4	0.0000

It can be seen from table 3 and table 4, we can establish an entity fixed effects model at the significant level of one percent. In addition, we also weigh the important statistical indicators among the entity fixed effects model, time fixed effects model and time and entity fixed effects model. Finally, we choice the entity fixed effects model. From the specific issues of the paper, due to study the relationship between accidents losses and highway mileages using 31 regions of China, and in order to further reflect the differences of geographical characteristics in different regions, so the entity fixed effects model is more appropriate than other models.

In sum, we choice the panel data model with regional fixed effects and variable intercept. The estimated results are shown in the following equation (8).

$$
\begin{aligned}
\ln Loss_{it} = {} & 0.97 + \alpha_i^* + 0.17 \ln Mileage_{it} + 0.31 \ln Accidents_{it} + 1.32D \\
& (0.27) \qquad (2.62) \qquad (7.81) \qquad (4.11) \\
& -0.17 \ln Accidents_{it} \times D + 0.97 \ln Loss_{i,t-1} \\
& (-4.74) \qquad (24.99)
\end{aligned} \tag{8}
$$

$$R^2 = 0.9644 \quad DW = 1.82 \quad F-statistic = 188.14 \quad \text{Prob}(F\text{-}statistic) = 0.0000$$

As shown in equation (8), the term ($\ln Mileage_{it} \times D$) doesn't appear in estimation formula. This is because that the regression coefficient is not significant by adding the explanatory variables into the model. In addition, the regional cross-section fixed effects coefficients are shown in table 5.

Table 5 The Cross-section Fixed Effects Coefficients in 31regions of China

Region i	α_i^* Estimate	Region i	α_i^* Estimate
Beijing	-1.42	Hubei	0.25
Tianjin	0.84	Hunan	0.79
Hebei	0.98	Guangdong	2.54
Shanxi	2.25	Guangxi	-0.68
Inner Mongolia	0.55	Hainan	-1.38
Liaoning	0.06	Chongqing	-0.14
Jilin	0.67	Sichuan	1.49
Heilongjiang	0.71	Guizhou	-0.31
Shanghai	-3.88	Yunnan	-0.90
Jiangsu	-0.34	Tibet	-0.01
Zhejiang	0.62	Shanxi	1.88
Anhui	-0.11	Gansu	-0.52
Fujian	1.13	Qinghai	-2.21
Jiangxi	1.63	Ningxia	-2.35
Shandong	1.11	Xinjiang	-3.21
Henan	-0.04		

As can be seen from the above estimates, the goodness of fit in the regression equation reaches 96.44%, which reflects the fit very well. The t statistics of explanatory variables in the brackets of equation (8) also pass the test significantly. The DW statistic is close to 2, which means the residual terms (ε_{it}) with non-autocorrelation. The F statistic is 188.14, and accompanied probability is 0.0000, which explain the differences among different regions to impact the model's setting significantly. Therefore, it is reasonable to use the panel data model with regional fixed effects and variable intercept.

From the national average level, we can also think that there is a long-term equilibrium relationship between accident losses and highway mileages. The coefficient of elasticity with accident losses on highway mileages is 5.41($0.17/(1-0.97)=5.41$ [6]). i.e. if highway mileages increases by 1%, then accident losses will increase by 5.41%. Moreover, it can be seen from table 5, the fixed effects coefficients (α_i^*) of different regions are also significant differences. The result is consistent with the levels of economic development and the conditions of road traffic in different regions of China.

2.3.4 Analysis of Results

Compared with the analysis of time series data and cross-sectional data, the analysis of panel data using entity fixed effects model contains more the number of samples, and is more desirable for estimated effect of the model. The analysis of panel data not only can describe the overall general characteristic, but also can reflect the differences among different entities. Therefore, it can reflect the actual situation of research question more objectively. In this paper, we draw the conclusion that there is a significant positive correlation relationship between accident losses and highway mileages with both unit root test and co-integration test methods of panel data. These studies will provide a useful reference to consider the relationship between net premium of vehicle insurance and driving mileages in the next section.

3. Research Reference of Net Premium of Vehicle Insurance and Driving Mileages

Similar to the previous empirical results, if there is a significant positive correlation relationship between net premium of vehicle insurance and driving mileages, then study the relationship has particular benefits for the insured. It can provide an option to save net premium by reducing driving mileages and to choose optimal net premium level according their own characteristics. Moreover, it also has important theoretical significance and practical value for accuracy, adequacy and fairness of pricing for insurance companies. Combined with the econometric model of the given accident losses and highway mileages, and learned from overseas study of vehicle insurance costs and driving mileages, we will give some practical research ideas. These explorations will provide effective practice reference to consider driving mileages and to be introduced and developed into GLM pricing for actuaries in China's property insurance companies.

3.1 Access to driving odometer information of the insured

Since the late 1990s, non-life insurance practitioners have gradually changed the vehicle premium charges in Europe and United States and other developed countries. i.e. vehicle premiums were to be collected based on driving mileage, it was called pay as you drive. The research was initiated by Pascal Noel that research expert of PAYD products in the Brookings institute. Some summary of researches can refer Litman(2005), Ferreira & Minikel(2010), Victoria Transport Policy Institute(2011). The PAYD products need to use remote communication such as GPS, and other information transmission technology. At present, there are two different operation modes in practice of foreign insurance companies. The first mode is designed to install wireless devices in the insured vehicle by the pioneers of Progressive corporation in 1998, and to record driving odometer information such as when, how much mileage, how to driving, etc. and further to send the information to property insurance companies. The second mode is the operation mode of Milemeter corporation, which determine vehicle insurance rates according to ratemaking variables, such as age of the insured, region of insured vehicle,

[6] Due to rounding, calculate directly the formula is not equal to 5.41.

vehicle type, etc. As well as the mode can provide different range of mileages for the insured, so that the insured free select appropriate mileages range by their own situation. Moreover, it verifies mileages by true odometer in the expiration of insurance period. Intuitively, the first mode needs to invest high cost for property insurance companies. Therefore, the second mode has more reference to China's property insurance companies.

For vehicle insurance businesses of China's property insurance companies, the period of vehicle insurance products is relatively short, and usually is one year. When issued new policy or renewal policy, relevant staff can obtain odometer information of the insured within insurance period, and further record driving mileages to the database. The information has important practical value for actuaries to analyze the relationship of net premium of vehicle insurance and driving mileages, so as to add the driving mileages into GLM pricing. Therefore, we advocate and take the implementation program to obtain mileages information as soon as possible of property insurance companies. In addition, because access to this information takes some time, during the period, qualified companies can also be insured vehicle to be classified by region, people and types of vehicle, design a questionnaire survey or telephone inquiries to obtain mileage information. On this basis, under the new challenges of vehicle insurance pricing and underwriting capacity upgrade, property insurance companies can maintain the existing business through first design and develop the vehicle insurance business considering driving mileages. It maybe a useful attempt to seek the growth points of new business.

3.2 The Ideas and Methods of Study the Relationship of Net Premium and Mileages

3.2.1 The econometric model of net premium, driving mileages and claims frequency

Combined with research ideas of the second section, considered specific company, actuaries can establish an econometric model involving total insurance cost or total net premium (i.e. total losses cost)[7] with total driving mileages, total claims frequency and other explanatory variables based on all insured vehicles, so as to study the relationship between vehicle insurance net premium and driving mileages from the company's overall level. Moreover, actuaries can also create another econometric model involving insurance cost or net premium with average driving mileages, average claims frequency and other explanatory variables based on unit insured vehicle, which study the relationship from the company's average level. Regardless of based on overall level or based on average level, all two research ideas can calculate the corresponding elastic coefficient. It is noteworthy that, the derived result of the second section is that highway mileage increases by 1%, accident losses will increase by 5.41%. Therein, highway mileage is the length of highway transport routes of regional statistics, and it is an objective variable mileage, so it can't be directly considered as a specific mileage of underwriting vehicle or average mileage. Accident losses are converted into the amount of losses of regional statistics. Therefore, Quantitative analysis method used in the empirical analysis only provides a positive correlation relationship between highway mileages (as an objective indicator) and regional aggregate losses significantly. This relationship will be to provide a reference for the establishment of econometric model between vehicle insurance net premiums of property insurance company and driving mileages.

3.2.2 Overseas studies on vehicle insurance costs and driving mileages

[7] Total insurance cost of insured vehicles is different from total net premium. The former includes total losses cost, also includes additional costs and profits, however, the latter refers to total losses cost.

Due to driving mileages statistics has been in the database of insurance companies in Europe and the United States and other developed countries. As a research expert of Berkeley, university of California, Edlin(1998) studies vehicle insurance premium of unit mileage. The results show that the elastic coefficient of vehicle insurance costs to driving mileages is between 1.42 and 1.85. In other words, driving mileages increases by 1%, insurance costs will increase by 1.42% to 1.85%. Once there is mileage statistics in the database of property insurance companies in China, we can learn from the research idea to study the elastic coefficient of vehicle insurance costs to driving mileages, and compare China's actual situation with the results of Edlin.

3.2.3 Mileage-based GLM pricing in vehicle insurance

Once obtained the mileage statistics, actuaries can also study the relationship between vehicle insurance net premiums and driving mileages directly in the GLM framework. In other words, based on known information, such as vehicle types, vehicle uses, driving districts, driving mileages, gender, age of driver, traffic accident records and driving behavior and other factors, they can select appropriate ratemaking variables, such as types of vehicle, age of driver, respective district of vehicles, driving mileages, and further create GLM of loss frequency and loss severity based on selected ratemaking variables respectively. The research not only can study the significance of the relationship between vehicle insurance net premiums and driving mileages, but also can further determine vehicle insurance rates on this basis.

In summary, we provide three ideas and methods to study the relationship between vehicle insurance net premiums and driving mileages. These methods mean that property insurance companies should obtain driving mileages information gradually. Until obtained mature data, they can incorporate driving mileages into ratemaking variables directly and develop GLM pricing for vehicle insurance products.

4. Further Proposal for Mileage-Based Pricing

Further study of mileage-based pricing in vehicle insurance products of China can be subdivided into the following four aspects.

First of all, for further empirical analysis, study the impacts of mileage on vehicle insurance net premium at the company's level. We should further obtain the relevant data of the domestic insurance companies, as well as collect driving mileage and other information in different regions and vehicle models through a questionnaire survey, use elastic analysis methods to study and test the correlation between traffic accidents losses and mileages.

Second, for theoretical model, we can also model the relationship between traffic accidents losses and mileages using appropriate statistical models and methods. The alternative approaches are as follows.

(1) Incorporate a mileage impact factor into generalized linear model pricing of vehicle insurance, as well as consider a variety of extended model in GLM, such as generalized additive model(GAM), generalized linear mixed models(GLMM), etc.

(2) Consider the general regression model, modeling of the relationship between per unit of driving mileage and per unit of net premium in vehicle insurance. (such as per unit of mileage should reflect the differences among the vehicle, driver and regional; mainly concern the mileage explanatory variable; use rate relative to consider other factors; etc)

(3) Bayesian statistical analysis models and methods, such as a variety of Bayesian nonlinear models for forecasting insurance loss payments; actuarial credibility theory; etc.

(4) Compare the pros and cons of these statistical models, study the similarities and differences, commit to regulate these models from an intuitive, consistent framework. For example, incorporate credibility theory into the GLM framework using more rigorous statistical methods.

Third, consider the differences of business structure among different property insurance companies, provide some possible suggestions of mileage-based pricing in practice.

Finally, study the positive externalities of mileage-based pricing in vehicle insurance. The optional programs are as follows.

(1) Use scenarios generation technique, intelligent algorithms and a variety of simulation methods, such as the MCMC simulation, Bootstrap method, etc; simulation analysis of the available date from domestic and overseas vehicle insurance.

(2) From an economic perspective, study the marginal private benefits and marginal social benefits of insurance companies, further investigate and discuss the positive externalities of mileage-based pricing in vehicle insurance from insurance, energy, transport and environment perspectives, as well as promote the insurance industry to participate in social management innovation.

(3) A Comparative study between mileage-based pricing model and traditional model of vehicle insurance.

5. Innovations and Conclusions

Since there is no mileage statistics in the database of almost all property insurance companies in China currently, the paper proposes to study the impact of driving mileages on net premiums of vehicle insurance innovatively. The paper makes an empirical analysis to study the long-term stable equilibrium relationship between the accident losses and highway mileages by using econometric methods of panel data. The result shows that there is a significant positive correlation relationship between the accident losses and highway mileages. These studies provide useful support and reference to consider the relationship between net premiums of vehicle insurance and driving mileages, and then make use of driving mileages to adjust net premiums of vehicle insurance within reasonable elastic range. In addition, the results of the paper have important theoretical significance and practical value for driving mileages to be introduced and developed into GLM pricing in China's vehicle insurance. As a further research direction, the study is intended to build a pricing model based on mileage of vehicle insurance products, so as to promote the independent development of intellectual property rights of mileage-based pricing in vehicle insurance products, upgrade of product structures, promote the reform of the commercial vehicle insurance premium management system of the China Insurance Regulatory Commission.

In the empirical analysis of the paper, four kinds of data processing and statistical analysis software, such as Excel, R software, Eviews7.0, Stata12.0 are used. Therein, these three-dimensional maps of figure 1 and figure 2 are drawn using R software; figure 3 shows that the trends of model variables containing three vertical axes over time using Excel; unit root test and co-integration test of panel data, model setting and parameter estimation results are obtained by using two econometric analysis software both Evies7.0 and Stata12.0. The outputs are basically the same. On the basis, we draw the main conclusions of the paper, i.e. highway mileages increase by 1%, accident losses will increase by 5.41%.

Compared with traditional pricing of vehicle insurance products without considering driving mileages, the proposed ideas and methods of vehicle insurance pricing in the paper have obvious advantages. These advantages are mainly reflected in the following four aspects. Firstly, the pricing method considering driving mileages has transmitted the loss cost of one kilometer more to the insured, making the insured may choose the optimal driving mileages to save premium, according to their own situation. Compared with the existing imposed fuel tax, vehicle purchase tax, environmental tax, fuel economy standards, traffic control and other external mandatory measures, the pricing method considering driving mileages provides am inherent power of financial incentive for the insured. It can be regarded as a new endogenous economic means. Secondly, the driving mileages of low-income families are less than of high-income families generally. Distinguished the insured by driving mileages not only can avoid cross-subsidies, but also conducive to realize social justice. In addition, it can also attract those uninsured low-income families to buy vehicle insurance products. Thirdly, for property insurance companies, on the one hand, the pricing method can promote actuaries to more accurately determine rates of vehicle insurance, and reasonably reflect the insurance cost for each insured so as to charged premiums more equitable. On the other hand, actuaries can have a quantitative analysis of each insured's driving conditions, in order to reduce the potential risk of loss and increase the company's underwriting profit. Fourthly, from the perspective of society as a whole, reducing vehicle's annual average mileages can decrease traffic congestion effectively, reduce vehicle collisions and insurance payments, decrease environmental pollution and greenhouse gas emission, enhance energy security and other comprehensive income. For some extent, it also help to establish a good public image of insurance companies in the low-carbon economic environment, and more conducive to the healthy and orderly development of insurance market.

References

[1] Brown, R.L. Minimum Bias with Generalized Linear Models[J]. Proceedings of the Casualty Actuarial Society, 1998, 75(143):187-217.

[2] Choi In. Unit root tests for panel data[J]. Journal of International Money and Finance, 2001, (20): 249-272.

[3] de Jong, P., Heller, G.Z. Generalized Linear Models for Insurance Data[M]. Cambridge: Cambridge University Press, 2008 : 64-80.

[4] Edlin, A.S. Per-Mile Premiums for Auto Insurance[R]. NBER working paper series, http://www.nber.org/papers/w6934, 1998, (1):1-40.

[5] Feldblum, S., Brosius, E. The Minimum Bias Procedures: A Practitioner's Guide[J]. Casualty Actuarial Society Forum, 2002, (1) :591-683.

[6] Ferreira, J.Jr., Minikel, E. Pay-As-You-Drive Auto Insurance In Massachusetts: A Risk Assessment And Report On Consumer, Industry And Environmental Benefits[R]. by the Department of Urban Studies and Planning, Massachusetts Institute of Technology, 2010.

[7] FU Lu-yang, Wu, C.S. General Iteration Algorithm for Classification Ratemaking[J]. Variance, 2007, 1(2):193-213.

[8] Im, K.S., Peasaran, M.H., Shin, Y. Testing for unit roots in heterogeneous panels[J]. Journal of Econometrics, 2003, 115: 53-74.

[9] Levin, A., Lin, C. F., Chu, C. Unit root test in panel data: asymptotic and finite-sample properties[J]. Journal of Econometrics, 2002, 108: 1-24.

[10] Litman, T. Pay-As-You-Drive Vehicle Insurance–Implementation, Benefits and Costs[R]. Victoria Transport Policy Institute, 2005(11) :1-13.

[11] Maddala, Wu, G.S. A comparative study of unit root tests with panel data and a new simple test[J]. Oxford Bulletin of Economics and Statistics, 1999, 61: 631-652.

[12] McCoskey, S., Kao, C. A residual-based test of the null of cointegration in panel data[J]. Econometric Reviews, 1998, 17: 57-84.

[13] Mildenhall, S.J. A Systematic Relationship between Minimum Bias and Generalized Linear Models[J]. Proceedings of the Casualty Actuarial Society 1999, 86, 393-487.

[14] Nelder, J.A., Wedderburn, R.W.M. Generalized Linear Models[J]. Journal of the Royal Statistical Society, Series A, 1972, 135(3): 370-384.

[15] Ohlsson, E., Johansson, B. Non-life Insurance Pricing with Generalized Linear Models[M]. European Actuarial Academy Series–Textbook, Springer-Verlag Berlin Heidelberg, 2010 : 39-135.

[16] Pedroni, P. Critical values for cointegration tests in heterogeneous panels with multiple regressors[J]. Oxford Bulletin of Economics and Statistics, 1999, 61: 653-678.

[17] Schmidt, K.D, Wünsche, A. Chain ladder, marginal sum and maximum likelihood estimation[J]. Blätter DGVM 1998, 23: 267-277.

[18] Victoria Transport Policy Institute. Pay-As-You-Drive Vehicle Insurance–Converting Vehicle Insurance Premiums Into Use-Based Charges[R]. TDM Encyclopedia, 2011(6).