Risk Classification and Claim Prediction: An Empirical Analysis from Vehicle Damage Insurance in Taiwan

Chu-Shiu Li
Department of Risk Management and Insurance
National Kaohsiung First University of Science and Technology
Kaohsiung, Taiwan

Chwen-Chi Liu
Department of Risk Management and Insurance
Feng Chia University
Taichung, Taiwan

Sheng-Chang Peng*
Department of Risk Management and Insurance
Ming Chuan University
Taipei, Taiwan

* Corresponding author, email: scpeng@mail.mcu.edu.tw.

Risk Classification and Claim Prediction: An Empirical Analysis from Vehicle Damage Insurance in Taiwan

Abstract

By conducting prediction models by logistic regression, this paper uses a unique vehicle insurance data set in Taiwan to examine whether rating characteristics are still effective under a bonus-malus system and to investigate whether extra information can help predict claim occurrences for vehicle damage insurance. The empirical results show that all current rating characteristics for vehicle damage insurance are significant factors to predict claim occurrence. Among all, claim coefficient, car age, and car model are relative important information for risk classification. In addition, we consider some extra information regarding both policyholder and automobile third-party liability insurance coverage to predict claim occurrence. As stated by the accuracy of prediction models, we find that claim record in the previous policy year is a useful information for risk classification.

Key Words: risk classification, claim prediction, automobile insurance

1. Introduction

Risk classification is an essential task in the insurance field from both theoretical and practical views. Unable to acquire complete information of their insureds, insurance companies determine premium rates through categorizing based on available information. Not only does risk classification closely relate to the efficiency of insurance market equilibrium, but it also functions as a necessary process for insurers to maintain solvency and fairness.

In the case of automobile insurance, it is common for insurers to use a number of a priori classification variables such as the main driver's age, gender, and occupation, and vehicle's usage and type, to adequately and fairly differentiate risk levels among policyholders (Lemaire, 1995). However, other characteristics such as driving behavior and driving history are also important in pricing automobile insurance rates. To include driving experience into risk classification, the bonus-malus system (BMS) or merit rating system has been widely used for a long time. A BMS rewards policyholders without filing any claims by providing a discount (or bonus) and penalizes policyholders involved in one or more accidents by adding extra premium (or malus).

There are usually two effects when a BMS is adopted by insurers: (i) it may prompt the insureds to drive carefully and reduce accident occurrences, and (ii) it may link driving risk to premiums more adequately. Utilizing automobile insurance data before and after Tunisia introduced a BMS to its vehicle insurance pricing system, Dionne and Ghali (2005) find a reduced probability of making claims for policyholders who remained with the same insurance company during the observed period. Moreno et al. (2006) provide a theoretical model to prove a bonus-malus contract can help eliminate fraud. Furthermore, a BMS might induce some policyholders not to report small claims (Lemaire, 1977) or to accumulate small losses to file one claim (Li et al., 2012) in order to avoid higher future premiums. Therefore, the effects of a BMS on both a policyholder's long-term driving behavior and subsequent claim behavior. For instance, when an insured has made a claim due to a car accident in the previous year, this year, he or she will drive more carefully to avoid filing a claims and paying higher premiums in the following year so that he or she can get a discount on future premiums.

We attempt to study the efficiency of pricing characteristics for risk classification in this paper through analyzing vehicle damage insurance data in Taiwan. In other words, we would like to investigate which characteristics currently applied by insurers significantly impact the occurrence of auto insurance claims. Especially, drivers' claim behavior might be affected by the BMS. Then it becomes a new issue that whether those characteristics for risk classification are still operative. Furthermore, we also want to examine whether some additional information which insurers are able to get from insureds could improve risk classification for vehicle damage insurance.

In regard of examining the characteristics of risk classification, it is common to test the relationships between a priori variables of risk classification and occurrence of ex post claims in order to identify which variable can differentiate risk levels of insureds. An effective risk classification variable, however, should have the ability to predict claim occurrence. It would be a useful way to examine the characteristics of risk classification from the results of prediction risk classification characteristics from analyzing prediction results. In other words, through setting different models to predict claim occurrence, we could investigate how different rating characteristics affect the following policy year. Particularly, we could estimate the importance of each different rating characteristic from comparing the accuracy of prediction models. As insurers develop their risk classification method for automobile insurance, our results might provide some ideas for them to underwrite policies or review their current pricing strategy.

Through conducting prediction models by logistic regression, this paper uses a vehicle insurance data set to examine whether rating characteristics are still effective under a BMS and to investigate whether extra information can help predict claim occurrences for vehicle damage insurance. Our empirical results show that all characteristics in the current rating system are significant factors to predict claim occurrence. Among all rating characteristics, claim coefficient, car age, and car model are relative important information for risk classification. For checking useful factors to progress in risk classification, we consider some additional information regarding both policyholder and automobile third-party liability insurance coverage. As stated by the prediction accuracy, claim record in the previous policy year is a useful information for risk classification.

This paper is organized as follows. In Section 2, we review literature on risk classification for automobile insurance. Then we introduce rating characteristics in Taiwan's vehicle damage insurance market in the next section. As for Section 4, we describe our empirical data and methodology. Our estimated models and prediction results will be presented in Section 5, and Section 6 concludes this paper.

2. Literature review

On the literature of risk classification, early researches focused on theoretical approaches, the efficiency of risk classification and its influence on social benefits, such as Hoy (1982), Crocker and Snow (1985, 1986, and 1992), Bond and Crocker (1991), etc. Crocker and Snow (2000) later reviewed previous theoretical views and summarized that risk classification can raise the efficiency of insurance markets under information asymmetry in a condition determined by each market's information system and its insureds' understanding of classification characteristics. Researches in this genre do not target at analyzing particular insurance types but discussing issues of risk classification based on dichotomy instead – studying cases of both perfect information and information asymmetry and of both competitive and non-competitive markets.

As for risk classification of vehicle insurance, it refers to questions oriented to practices. Each country's automobile insurance system often conducts risk classification through characteristics of vehicle – car usage, brand, and style, and characteristics of policyholder – insured’s gender, age, and claim record (Lemaire, 1995). Between these two kinds of classification characteristics, there are little argument on characteristics of vehicle and accordingly scant discussion. On the contrary, there is plenty discussion on characteristics of policyholder. The main reason behind this result is that European and American countries used to argue over whether discrimination rates should be decided by gender, age, or race, such unalterable characteristics because discrimination cases took place. Therefore, researchers often investigate vehicle insurance data by analyzing examples applying gender and age as classification variables. For example, Butler et al. (1988) argued that pricing rules using gender to regulate discrimination rates is questionable. They applied actual data to analyze car accident records for both male and female drivers, then compared the differences between collected premiums and claim losses from insurance companies, and found that insured US females were overcharged for their vehicle insurance.

Puelz and Kemmsies (1993) used data of three personal vehicle insurance policies in Georgia, USA, including vehicle collision coverage, full coverage, and liability coverage, to evaluate how gender and other demographic variables impact on premium pricing. Their empirical research results showed that gender significantly affects premium rates, yet its influential degree is relatively less than other variables such as driving record, age, location, and vehicle type. Accordingly, it might be unnecessary for supervising administration to spend much time making laws against insurance rate pricing resorts to gender. Relying on various existing viewpoints on restricting this kind of risk classification, Harrington and Doerpinghaus (1993) examined administrative regulations of risk classification in automobile insurance market. They concluded their research by pointing out that although the effects of administrative measures are unclear; they indirectly lead to risk cost or claim control cost increases as they distort claim incentives. As a result, both the supply and the coverage of individual insurance might be impacted.

Until present day, scholars often propose related analyses of similar car insurance pricing characteristics. A recent research conducted by Doerpinghaus et al. (2008) applied the Closed Claim Survey data of year 1997 provided by the Insurance Research Council (IRC) in USA to study relations between claim filers' demographic characteristics and claim payment of third party liability vehicle insurance policies. Via three possible economic theories discussing diverse risk attitudes and differences and discrimination on negotiation costs, this study explained how various demographic characteristics result in different claim amounts and set empirical models to demonstrate possible relations between demographic characteristics and claim amount. The results indicated that while controlling other variables, female insureds receive less claim payments than male ones and married insureds receive more claim payments than single ones. The relation between age difference and claim payment is insignificant in Doerpinghaus' study, but there are other researches arguing over age. Brown et al. (2007) mentioned previous argument in Canada around whether to include age as a vehicle insurance pricing characteristic, and six among ten provinces refuse to do so. Nevertheless, other documented studies which applied car accident data have found age's influence on car accident occurrence. For instance, Braver and Tempel (2004) and Tefft (2008) identified higher accident tendencies for young and elder drivers. To put their findings into a figure with car accident loss versus age, then the figure shows a line close to a U shape. Such results respond to the rate regulations in practice in Taiwan, which apply higher rate coefficients upon young and elder insureds.

In addition, several papers investigate the effectiveness of marriage as a vehicle insurance risk classification characteristic because similar to the cases of gender and age, there is dispute over whether taking marital status into account is a kind of discrimination. Gardner and Marlett (2007) retrospected the history of US vehicle insurance market to envision future trends on vehicle insurance coverage, rates, market management, and related laws. On the subject of estimating rates according to one's marital status, the most common reason to support this application is marital status' relativity, considering that married drivers are calmer and more responsible than unmarried drivers. Meanwhile, opposing opinions emphasize that not every driver is granted the right of marriage; for example, forty-five states in USA forbid same sex marriage, so only heterosexual drivers are qualified for marital status rate discount. Moreover, because less people orient to marital relationships and the (average) age at first marriage rise at present day, the effects of marital status discount will become less critical to insurance purchasers. While most insurance purchasers do not agree that marital status should be an insurance rate pricing characteristic, the empirical analysis by Doerpinghaus et al. (2008) has proved that marital status impacts on the total of claim payment.

On the other hand, researches target on claim experience as a risk classification characteristic do not take place until recent years. Even though there were early practical cases of claim experience, most countries adopted this characteristic rather late. Dionne and Ghali (2005) used the vehicle insurance data of Tunisia before and after it adopted a BMS and found that the probability for insureds who stayed with the same insurance company decreases. Also, Moreno et al. (2006) designed theoretical models which indicated that BMSs can prevent insurance fraud. The research literature above suggests that applying BMSs induce insureds to lower driving hazards, but BMSs may change insureds' claim behavior, encouraging bonus hunger (Lemaire, 1977). Because of the linkage between premium rates and filed claim numbers, insureds may hesitate to report small claims or accumulate small losses to file one claim before policies mature (Li et al., 2013) in order to avoid increasing their claim totals.

Except for analyses on those common risk classification characteristics above, there is also plenty discussion with other factors. Via statistical analysis, Kellison et al. (2003) examined the relationships between policyholder's credit record and policies with claim record, and they identified that those with worse credit scores report greater losses (including both loss frequency and loss degree), which further proved that credit record provides information unavailable in the current underwriting system. A related yet slightly different research by Miller and Smith (2003) analyzed six types of private car insurance policies and insurance scores only applicable to this kind of vehicle insurance. However, these two papers did not explain how credit record assists insurance companies in risk evaluation. To bridge this gap, Brockett and Golden (2007) first reviewed related literature before investigating the relations between credit record and car insurance loss from their studying of biological, psychological, and behavioral attributes and financial assumption of risk regarding these attributes. They concluded that credit evaluation could be turned into useful underwriting information only when individual biological and psychological differences are reflected upon the loss risks of insured vehicles.

Besides, Bair et al. (2012) predicted car accident occurrence based on vehicle maintenance record through data of compulsory vehicle liability insurance and the unique maintenance record in Taiwan. They traced lower accidence occurrence probability from insured cars which follow their maintenance schedules, but they did not find significant effects on loss degree. Accordingly, Li et al. (2013) analyzed vehicle insurance data in Taiwan, and they found a significantly less probability of filing vehicle physical damage claims for insureds who purchase new cars along with vehicle physical damage policies that bundle with high insured amount voluntary vehicle liability policies. To put it in another way, the purchasing behavior of bundled insurance coverage also generates useful risk evaluation information.

3. Rating Characteristics for Vehicle Damage Insurance in Taiwan

In Taiwan's property insurance market, vehicle insurance is the primary business source for insurance providers. Based on previous statistical data, the income from automobile insurance premium accounts for approximately fifty per cent of all premium income; meanwhile, vehicle insurance claims comprise sixty per cent of the claim total. Judging from both revenue and expense aspects, steadily managing automobile insurance business or not is critical to each insurance company's development.

For a long time, the Taiwanese automobile insurance market was regulated, so premium rates were calculated and announced by supervising administration. Officials decided basic premium for each insurance policy and related characteristics which determined discrimination rates. Under rate regulations, the pricing characteristics of vehicle physical damage insurance resorted to characteristics of policyholder and characteristics of insured vehicle, while characteristics of policyholder take insured's age and gender (gender-age coefficient) and claim record (claim coefficient) into account. For detailed information on gender-age coefficient, please see Table 1. In general, the coefficient for male is higher than female, and young drivers are also noted with relatively higher coefficients. The claim coefficient is calculated based on the claim record in previous three years estimated as cumulative claim points. As for characteristics of insured vehicle, official rates consider insured vehicle's usage, type, age, brand, and style (manufacture coefficient). Therefore, the administration calculates premium of vehicle physical damage insurance with this following equation:

$$\text{Premium} = \text{Basic Premium} \times (\text{Gender-age Coefficient} + \text{Claim Coefficient}) \times \text{Manufacture Coefficient}$$

Following the global trend of rate liberalization, insurance companies in Taiwan can determine their own vehicle insurance premium rates since April 2009. However, a great number of insurance providers stick to the official rates announced by administrative institutions.

[Insert Table 1]

4. Data and Methodology

Data

This paper uses a data set of private vehicle damage insurance policy and claim information for the policy years between 2010 and 2012 from the Taiwan Insurance Institute. The policy information includes demographic characteristics of the policyholder (age, gender, and marital status), characteristics of the vehicle (car age, car model, and exhaust), premium, and deductible type. The claim information includes demographic characteristics of the claimed driver (age, gender, and marital status), claim date, claim payment, and cause of accident.

The vehicle damage insurance policy covers accidents to the car, including rollover, lightning, fire, explosion, damage from flying objects, and collision, and it contains a deductible option. Policyholders can choose policies with or without a deductible. There are two types of deductibles: increasing per-claim deductible (3,000/5,000/7,000 New Taiwan Dollar) and straight deductible. In our data set, about 11 percent of policyholders purchase a vehicle damage insurance policy with a deductible, and 81 percent of them choose increasing per-claim deductible. Since different types of deductibles will affect whether an insured makes a claims for car accidents, we exclude those policies with straight deductibles to avoid heterogeneity bias from deductible types.

[Insert Table 2]

Our data are reorganized by policy year from 2010 to 2012. The number of policies and claim ratio are shown in Table 2. Depending on claim information, all proportions (claim ratios) of claimed policy numbers in all policies by policy year are about 48 percent. Moreover, we also observe the claim coefficient, which can serve as the long-term claim history. They are all negative in average for three policy years, which shows that many policyholders had a discount for vehicle damage insurance premium.

[Insert Figure 1]

As shown in Figure 1, the shares of new cars exceed 40 percent, and one-year cars are about 20 percent. The shares of other car ages display a declining tendency. Accordingly, we observe the claim history of both new car and old car policies, shown in Table 2. New car policies have higher claim ratios than old car policies for three policy years. Similarly, compared with renewed policies, there are relative high claim ratios for new policies which include new car policies and those policies transferred from other insurers. For insurance companies, renewed policyholders seem have lower risk. The claim coefficient displays consistent results mentioned above.

Methodology

This paper introduces logistic regression to conduct prediction models. After estimating coefficients of predictors, we proceed with estimation and holdout samples for prediction models. Depending on our data set, we use samples in the 2010 and 2011 policy years to estimate prediction models and make in-sample predictions, respectively. Subsequently, for out-sample predictions, samples in 2011 and 2012 are respectively used (correspondingly applied). Therefore, we have two sub data sets of 2010-2011 and 2011-2012 policy years to make estimations and predictions to confirm for predictive consistency. For examining the efficiency of prediction characteristics for claim occurrence, the model is set as follow:

$$\Pr(Claim = 1 | X_1, X_2, X_3) = \mathrm{F}(X_1\beta_1 + X_2\beta_2 + X_3\beta_3) \tag{1}$$

where Claim is a binary variable for whether a claim was filed during the insurance period; X_1 are demographic characteristics of the policyholder (such as age, gender, and marital status); X_2 are characteristics of the vehicle (such as car age, exhaust, car model, and domestic car); X_3 are other variables (such as insured district, insurance company, claim experience in the last policy year, and so on); β_1, β_2, and β_3 are estimated coefficient vectors.

Prediction accuracy by the logistic regression model can be examined by a classification table. A comparison of the predicted probability with the cutpoint can distinguish whether a predicted event will occur. If the probability of the predicted event is greater than or equal to the cutpoint, this defines that the predicted event will occur; otherwise it will not occur. The classification table applied in our study is shown in Table 2. Of total claimed policies (A+B), A is the number of policies that the predicted probability correctly forecast a claim filed during the policy period, and B is the number of policies falsely predicted. The sensitivity (= A / (A+B)) is the percentage of prediction accuracy of total claimed policies. Similarly, the specificity (= D / (C+D)) is the percentage of prediction accuracy of total non-claimed policies (C+D). Total correct (= (A+D) / (E+F)) is the percentage of prediction accuracy of total policies. E and F are numbers of claimed and non-claimed policies, respectively. The cutpoint is usually 0.5, but this paper would like to use a sensitivity analysis to determine it through the highest percentage of prediction accuracy from total policies.

[Insert Table 3]

5. Empirical Results

After controlling for characteristics of the insured and vehicle, we predict the claim probability using prediction models and then determine whether a claim is made during the policy period by the cutpoint which is determined by sensitivity analysis. We apply the percentages of correct prediction to the estimation sample (in the estimated policy year) and holdout sample (in the next policy year) in order to evaluate the accuracy of the prediction models. Prediction results are divided into the 2010 and 2011 policy years. We only display prediction results for the 2010 policy year, and others are shown in Appendix.

Firstly, we examine individual effect of rating characteristics, and results for the 2010 policy year are shown in Table 4. From estimation results, all rating factors, which include insured age, gender, claim coefficient, car age, exhaust, and car model, have significant effects on claim occurrence via Wald test, shown in Model 1. Then we proceed prediction for claim occurrence through individually excluding one of six rating characteristics from basic model, Model 1, and the results are shown in Models 2-7. When claim coefficient is removed out of the prediction model, the ratio of total correct in Model 3 is slightly smaller than it in Model 1 for estimation sample, and the difference rises for holdout sample. We have similar findings for car age and car model. These results demonstrate that claim coefficient, car age, and car model are relative important among current rating characteristics. There are consistent results in the 2011 policy year, shown in Appendix A.

Secondly, we implement claim predictions for including additional information, shown in Table 5. Model 1 is basic model which contains all factors in rating system. Models 2-6 individually join additional information into basic model, including insured's married status, claim filed in the last policy year, liability coverage for third-party's body and property, and claim coefficient of liability coverage. In Model 3, we find that the information of claim record in the previous policy year could improve prediction accuracy. Even if we contain all additional information in the prediction model (Model 6), the ratios of total correct for estimation and holdout sample are similar with those in Model 3. This finding exhibits that claim history in the last policy year is an important information for improving risk classification. Other additional information, however, cannot have consistent and robust estimation results, and also cannot improve prediction accuracy. We have consistent outcomes in the 2011 policy year, shown in Appendix B.

According to significantly different claim ratios among new car policies, new policies, and renewed policies, this paper also proceeds prediction analyses for three sub-samples, shown in Table 5. In Panels A and B, there are similar prediction results. The ratios of sensitivity (prediction accuracy of claimed policies) are all higher than 70 percent for estimation and holdout samples. On the contrary, the ratios of specificity (prediction accuracy of non-claimed policies) are lower than 60 percent. As insurers cannot get more claim information for new car policies and new policies, other additional information cannot improve prediction accuracy as mentioned above. As for renewed policies, the ratios of sensitivity (prediction accuracy of claimed policies) are about 40 percent, and the ratios of specificity (prediction accuracy of non-claimed policies) are higher than 80 percent. These prediction results of renewed policies are greatly different compared with new policies or new car policies. From those ratios of total correct in Models 2-6, we can have the same finding that claim history in the last policy year is an important information for improving risk classification.

6. Conclusions

A driver's past claim history has been considered as one of the most important variable to predict the future number of claims. However, not only can a BMS more adequately link driving risk to premiums, it can also prompt the insured to drive more carefully and reduce accident occurrences. As policyholders might change their claim behavior due to a BMS, it is interesting to examine whether rating characteristics are still effective. In addition, this paper uses an automobile insurance data set to investigate whether extra information can help predict claim occurrences and improve risk classification for vehicle damage insurance. We conduct prediction models by logistic regression. The percentages of correct prediction for estimation and holdout samples are used to evaluate the accuracy of prediction models.

The empirical results show that all characteristics in the current rating system are significant factors in prediction models, and claim coefficient, car age, and car model are relative important information for risk classification. As stated by the highest prediction accuracy, the predictor of claim record in the previous policy year is more helpful than other extra information among all prediction models. From the sub-sample of renewed policies, we have a consistent and robust result.

Reference

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

Bair, Shyi-Tarn, Rachel J. Huang, and Kili C.Wang, 2012, “Can Vehicle Maintenance Records Predict Automobile Accidents?” *Journal of Risk and Insurance*, 79(2): 567-584.

Bond, Eric W. and Keith J. Crocker, 1991, “Smoking, Skydiving and Knitting: The Endogenous Categorization of Risks in Insurance Markets with Asymmetric Information,” *Journal of Political Economy*, 99: 177-200.

Braver, E. R. and R. E. Trempel, 2004, “Are Older Drivers Actually at Higher Risk of Involvement in Collisions Resulting in Deaths or Non-Fatal Injuries among Their Passengers and Other Road Users?” *Injury Prevention*, 10: 27–32.

Brockett, P. L., and L. L. Golden, 2007, “Biological and Psychobehavioral Correlates of Credit Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works,” *Journal of Risk and Insurance*, 74(1): 23-63.

Brown, R. L., D. Charters, S. Gunz, and N. Haddow (2007), “Colliding Interests－Age as an Automobile Insurance Rating Variable: Equitable Rate-Making or Unfair Discrimination?” *Journal of Business Ethics*, 72: 103-114.

Butler, P., T. Butler, and L. Williams, 1988, “Sex-Divided Mileage, Accident, and Insurance Data Show that Auto Insurers Overcharge Most Women,” *Journal of Insurance Regulation*,” 6: 243-284, 373-416.

Crocker, Keith J. and Arthur Snow, 1985, “The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information,” *Journal of Public Economics*, 26: 201-219.

Crocker, Keith J. and Arthur Snow, 1986, “The Efficiency Effects of Categorical Discrimination in the Insurance Industry,” *Journal of Political Economy*, 94: 321-344.

Crocker, Keith J. and Arthur Snow, 1992, "The Social Value of Hidden Information in Adverse Selection Economies," *Journal of Public Economics*, 48: 317-347.

Crocker, K. and A. Snow, 2000, "The Theory of Risk Classification," in G. Dionne (ed.), *Handbook of Insurance*, 245–276, London: Kluwer.

Dionne, G. and O. Ghali, 2005, "The (1992) Bonus-Malus System in Tunisia: An Empirical Evaluation," *Journal of Risk and Insurance*, 72: 609-633.

Doerpinghaus, H., J. Schmit and J. J-H. Yeh, 2008, "Age and Gender Effects on Auto Liability Insurance Payouts," *Journal of Risk and Insurance*, 75(3): 527-550.

Harrington, S. and H. Doerpinghaus, 1993, "The Economics and Politics of Automobile Insurance Rate Classification," *Journal of Risk and Insurance*, 60(1): 59-84.

Hoy, Michael, 1982, "Categorizing Risks in the Insurance Industry," *Quarterly Journal of Economics*, 97: 321-336.

Hoy, Michael, Chu-Shiu Li, Chwen-Chi Liu and Sheng-Chang Peng, 2012, "Risk Classification, Rating Innovation, and Multiple Contracts in Automobile Insurance Market," 39th Seminar of the European Group of Risk and Insurance Economists, Palma de Mallorca, Spain.

Iversen, H. and T. Rundmo, 2002, "Personality, Risky Driving and Accident Involvement among Norwegian Drivers, Personality and Individual Differences," 33: 1251-1263.

Kellison, B., P. Brockett, S.-H. Shin, and S. Li, 2003, "A Statistical Analysis of the Relationship Between Credit History and Insurance Losses," Bureau of Business Research: McCombs School of Business, The University of Texas at Austin.

Lemaire, J., 1977, La soif du bonus, *Astin Bulletin*, 9: 181-190.

Lemaire, J., 1985, *Automobile Insurance: Actuarial Models*, Boston, MA: Kluwer Academic Publishers.

Lemaire, J., 1995, *Bonus-Malus Systems in Automobile Insurance*, Boston, MA: Kluwer Academic Publishers.

Li, Chu-Shiu, Chih-Hao Lin, Chwen-Chi Liu, and A. Woodside, 2012, "Dynamic Pricing in Regulated Insurance Markets with Heterogeneous Insurers: Strategies Nice versus Nasty for Customers," *Journal of Business Research*, 65: 968-976.

Li, Chu-Shiu, Chwen-Chi Liu, and Sheng-Chang Peng, 2012, "Is Claim Experience a Good Factor to Predict Automobile Insurance Claim Occurrences?" Asia-Pacific Risk and Insurance Association 16th Annual Conference, Seoul, South Korea.

Li, Chu-Shiu, Chwen-Chi Liu, and Sheng-Chang Peng, 2013, "The Expiration Date Effects of Automobile Insurance Contracts: The Curious Case of Last Policy Month Claims in Taiwan," *Geneva Risk and Insurance Review*, 38: 23-47.

Li, Chu-Shiu, Chwen-Chi Liu, and Sheng-Chang Peng, 2013, "Bundled Automobile Insurance Coverage and Accidents," *Accident Analysis and Prevention*, 50: 64-72.

Li, Chu-Shiu, Chwen-Chi Liu, and Jia-Hsing Yeh, 2007, "The Incentive Effects of Increasing Per-Claim Deductible Contracts in Automobile Insurance," *Journal of Risk and Insurance*, 74: 441-459.

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

Moreno, I., F Vázquez, and R. Watt, 2006, "Can Bonus-Malus Allieviate Insurance Fraud?" *Journal of Risk and Insurance*, 73: 123-151.

Puelz, R., and W. Kemmsies, 1993, "Implications for Unisex Statutes and Risk-pooling: The Costs of Gender and Underwriting Attributes in the Automobile Insurance Market," *Journal of Regulatory Economics*, 5(3): 289-301.

Tefft, Brian C., 2008, "Risks Older Drivers Pose to Themselves and to Other Road Users," *Journal of Safety Research*, 39(6): 577-582.

Wang, Jennifer L., Ching-Fan Chung, and Larry Y. Tzeng, 2008 "An Empirical Analysis of the Effects of Increasing Deductibles on Moral Hazard," *Journal of Risk and Insurance*, 75: 551-566.

Table 1. Gender-age coefficients

Age	Male	Female
Under 20	1.89	1.70
20 or above but under 25	1.74	1.57
25 or above but under 30	1.15	1.04
30 or above but under 60	1.00	0.90
60 or above but under 70	1.07	0.96
70 or above	1.07	0.96

Table 2. Claim history

	Number of policies			Claim ratio			Claim coefficient		
Policy year	2010	2011	2012	2010	2011	2012	2010	2011	2012
Total policies	135,082	141,809	130,932	48.09%	48.07%	48.70%	-0.15	-0.17	-0.17
New cars	61,236	62,963	49,957	59.13%	58.61%	60.80%	-0.01	-0.02	-0.03
Old cars	73,846	78,846	80,975	38.94%	39.64%	41.24%	-0.27	-0.29	-0.27
New policies	79,112	86,611	76,146	55.97%	55.67%	55.04%	-0.05	-0.06	-0.07
Renewed policies	55,970	55,198	54,786	36.96%	36.13%	39.88%	-0.31	-0.33	-0.33

Note: Claim ratio is the proportion of claimed policy numbers in all policies.

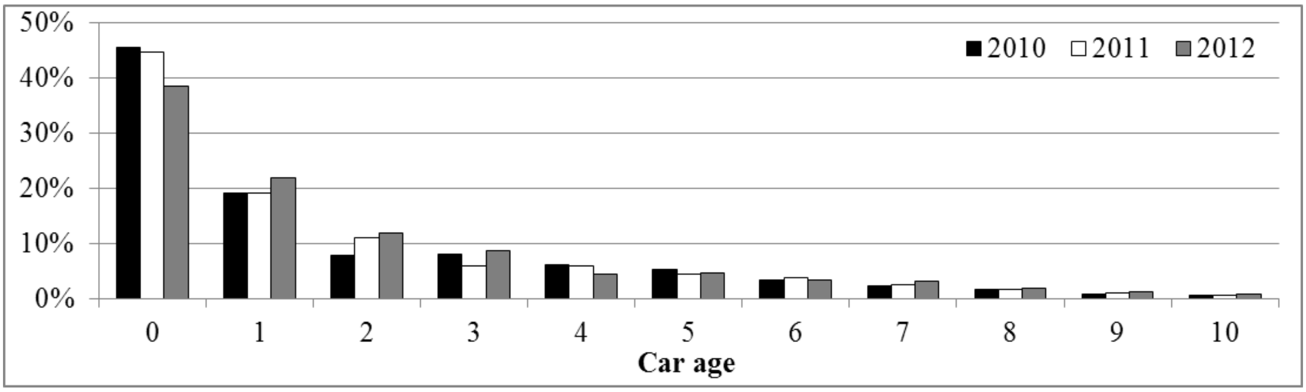

Figure 1. Car age pattern of the whole sample for each policy yea

Table 3. Classification table

Actual	%	Predicted: Claim	Predicted: No Claim
Claim	Sensitivity	A	B
No Claim	Specificity	C	D
	Total correct	E	F

Table 4. Examination results of individual effect of rating characteristics for the 2010 policy year

Predictor	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Wald tests of individual effects (Wald Chi-Square and P-value)							
Insured age	20.78		Y	Y	Y	Y	Y
	(0.0004)						
Male insured	17.53	Y		Y	Y	Y	Y
	(<.0001)						
Claim coefficient	1304.23	Y	Y		Y	Y	Y
	(<.0001)						
Car age	731.11	Y	Y	Y		Y	Y
	(<.0001)						
Exhaust	88.28	Y	Y	Y	Y		Y
	(<.0001)						
Car model	953.98	Y	Y	Y	Y	Y	
	(<.0001)						
Others	Deductible, Insured district, Insurance company						
Classification Accuracies for Estimation and Holdout Samples (%)							
Estimation sample							
Sensitivity	64.00	63.90	63.90	63.60	64.20	64.20	64.20
Specificity	71.50	71.50	71.60	71.40	71.10	71.30	70.60
Total Correct	67.90	67.90	67.90	67.70	67.80	67.90	67.50
Observations				135,082			
Holdout sample							
Sensitivity	64.20	64.18	64.21	65.37	63.79	64.37	64.86
Specificity	67.78	67.77	67.73	65.76	67.74	67.57	66.47
Total Correct	66.06	66.05	66.04	65.57	65.84	66.03	65.69
Observations				141,809			

Notes: The sensitivity is the percentage of prediction accuracy of the total claimed policies. The specificity is the percentage of prediction accuracy of the total non-claimed policies. Total correct is the percentage of prediction accuracy of the total policies.

Table 5. Estimation and prediction results of additional information for the 2010 policy year

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Logit Predictor Coefficients for Estimation Sample						
Basic risk factors	Y	Y	Y	Y	Y	Y
Married insured		0.043 ***				0.041 **
		(0.016)				(0.016)
Claim filed in last policy year			0.487 ***			0.486 ***
			(0.016)			(0.016)
Body liability coverage				-0.003		-0.004 *
				(0.002)		(0.002)
Property liability coverage				0.127 ***		-0.071 **
				(0.032)		(0.035)
Claim coefficient of liability coverage					-0.181 ***	-0.072
					(0.058)	(0.058)
Classification Accuracies for Estimation and Holdout Samples (%)						
Estimation sample						
Sensitivity	64.00	64.00	65.40	63.90	64.40	65.80
Specificity	71.50	71.50	71.30	71.50	71.20	71.10
Total Correct	67.90	67.90	68.50	67.90	67.90	68.50
Observations			135,082			
Holdout sample						
Sensitivity	64.20	64.24	65.44	64.25	63.44	64.42
Specificity	67.78	67.74	67.55	67.70	68.77	68.68
Total Correct	66.06	66.06	66.53	66.04	66.20	66.63
Observations			141,809			

Notes: ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels. Basic risk factors include a limited number of a priori classification variables for calculating premiums. The sensitivity is the percentage of prediction accuracy of the total claimed policies. The specificity is the percentage of prediction accuracy of the total non-claimed policies. Total correct is the percentage of prediction accuracy of the total policies.

Table 6. Prediction results of additional information for different sub-samples for the 2010 policy year

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Basic risk factors	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Married insured		Y				Y		Y				Y
Claim filed in last policy year			Y			Y			Y			Y
Body liability coverage				Y		Y				Y		Y
Property liability coverage				Y		Y				Y		Y
Claim coefficient of liability coverage					Y	Y					Y	Y
	Estimation sample						**Holdout sample**					
Panel A: New car policies												
Sensitivity	77.90	78.00		78.00	78.30	78.30	74.31	74.35		74.38	73.47	73.37
Specificity	55.10	55.30		55.20	54.50	54.50	53.45	53.37		53.35	55.04	55.27
Total Correct	68.60	68.70		68.70	68.70	68.70	65.68	65.67		65.68	65.84	65.87
Observations	61,236						62,963					
Panel B: New policies												
Sensitivity	77.00	77.00		77.00	77.60	78.00	71.38	71.39		71.39	70.33	70.25
Specificity	56.10	56.10		56.10	55.30	55.50	54.40	54.39		54.44	56.24	56.54
Total Correct	67.80	67.80		67.80	67.90	68.20	63.85	63.85		63.87	64.08	64.18
Observations	79,112						86,611					
Panel C: Renewed policies												
Sensitivity	40.50	40.50	41.00	40.70	41.00	41.50	43.15	43.23	44.83	43.50	43.29	44.71
Specificity	84.80	84.80	85.50	84.60	84.40	85.20	85.04	85.04	85.66	84.83	84.93	85.52
Total Correct	68.40	68.40	69.10	68.40	68.20	68.90	69.90	69.93	70.90	69.90	69.88	70.78
Observations	55,970						55,198					

Notes: *** denotes statistical significance at the 1 percent level. Basic risk factors include a limited number of a priori classification variables for calculating premiums. The sensitivity is the percentage of prediction accuracy of the total claimed policies. The specificity is the percentage of prediction accuracy of the total non-claimed policies. Total correct is the percentage of prediction accuracy of the total policies.

Appendix A. Examination results of individual effect of rating characteristics for the 2011 policy year

Predictor	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Wald tests of individual effects (Wald Chi-Square and P-value)							
Insured age	24.92 (<.0001)		Y	Y	Y	Y	Y
Male insured	42.83 (<.0001)	Y		Y	Y	Y	Y
Claim coefficient	1079.57 (<.0001)	Y	Y		Y	Y	Y
Car age	1121.49 (<.0001)	Y	Y	Y		Y	Y
Exhaust	100.47 (<.0001)	Y	Y	Y	Y		Y
Car model	1548.85 (<.0001)	Y	Y	Y	Y	Y	
Others	Deductible, Insured district, Insurance company						
Classification Accuracies for Estimation and Holdout Samples (%)							
Estimation sample							
Sensitivity	59.40	59.30	59.20	59.00	58.80	59.70	59.40
Specificity	73.40	73.50	73.50	73.20	73.30	73.00	71.80
Total Correct	66.70	66.60	66.60	66.30	66.40	66.60	65.90
Observations				141,809			
Holdout sample							
Sensitivity	52.01	52.00	51.96	52.07	52.17	51.89	52.42
Specificity	65.40	65.34	65.45	64.04	64.81	65.57	63.94
Total Correct	58.88	58.85	58.88	58.21	58.66	58.91	58.33
Observations				130,932			

Notes: The sensitivity is the percentage of prediction accuracy of the total claimed policies. The specificity is the percentage of prediction accuracy of the total non-claimed policies. Total correct is the percentage of prediction accuracy of the total policies.

Appendix B. Estimation and prediction results of additional information for the 2011 policy year

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Logit Predictor Coefficients for Estimation Sample						
Basic risk factors	Y	Y	Y	Y	Y	Y
Married insured		0.107 ***				0.084 ***
		(0.015)				(0.016)
Claim filed in last policy year			0.355 ***			0.397 ***
			(0.016)			(0.016)
Body liability coverage				0.036 ***		0.037 ***
				(0.001)		(0.001)
Property liability coverage				-0.456 ***		-0.569 ***
				(0.026)		(0.030)
Claim coefficient of liability coverage					-0.089	0.039
					(0.057)	(0.058)
Classification Accuracies for Estimation and Holdout Samples (%)						
Estimation sample						
Sensitivity	59.40	59.30	60.30	61.80	60.80	64.50
Specificity	73.40	73.50	73.50	72.20	72.60	71.40
Total Correct	66.70	66.60	67.20	67.20	66.80	68.00
Observations	141,809					
Holdout sample						
Sensitivity	52.01	52.08	53.06	58.51	51.24	58.64
Specificity	65.40	65.38	65.23	60.83	66.35	61.68
Total Correct	58.88	58.90	59.31	59.70	58.99	60.20
Observations	130,932					

Notes: *** denotes statistical significance at the 1 percent level. Basic risk factors include a limited number of a priori classification variables for calculating premiums. The sensitivity is the percentage of prediction accuracy of the total claimed policies. The specificity is the percentage of prediction accuracy of the total non-claimed policies. Total correct is the percentage of prediction accuracy of the total policies.

Appendix C. Prediction results of additional information for different sub-samples for the 2011 policy year

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Basic risk factors	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Married insured		Y				Y		Y				Y
Claim filed in last policy year			Y			Y			Y			Y
Body liability coverage				Y		Y				Y		Y
Property liability coverage				Y		Y				Y		Y
Claim coefficient of liability coverage					Y	Y					Y	Y
	Estimation sample						**Holdout sample**					
Panel A: New car policies												
Sensitivity	74.70	77.70		78.80	78.20	80.30	61.80	66.84		69.85	64.46	68.61
Specificity	54.10	51.20		52.50	51.10	51.20	47.61	44.67		39.96	47.24	41.73
Total Correct	66.20	66.80		67.90	67.20	68.50	56.24	58.15		58.13	57.71	58.07
Observations			62,963						49,957			
Panel B: New policies												
Sensitivity	80.20	79.80		77.00	81.20	78.70	73.04	73.11		71.69	71.75	70.54
Specificity	45.70	46.30		51.60	44.60	50.10	45.10	45.09		43.46	46.95	45.39
Total Correct	64.90	65.00		65.80	65.30	66.30	60.49	60.51		59.00	60.61	59.23
Observations			86,611						76,146			
Panel C: Renewed policies												
Sensitivity	47.20	47.50	48.50	48.40	48.50	50.20	33.08	33.08	34.05	36.91	33.00	38.40
Specificity	83.40	83.20	84.90	82.50	82.60	83.40	82.56	82.50	83.75	79.42	82.51	80.27
Total Correct	70.30	70.30	71.70	70.20	70.10	71.30	62.82	62.79	63.93	62.46	62.76	63.57
Observations			55,198						54,786			

Notes: Basic risk factors include a limited number of a priori classification variables for calculating premiums. The sensitivity is the percentage of prediction accuracy of the total claimed policies. The specificity is the percentage of prediction accuracy of the total non-claimed policies. Total correct is the percentage of prediction accuracy of the total policies.